

TranState Working Papers

POLITIKPOSITIONEN IM REFORMPROZESS
DES ÖFFENTLICHEN DIENSTES
ZUR ÜBERTRAGBARKEIT DER
QUANTITATIVEN TEXTANALYSE

KAMIL MARCINKIEWICZ
JENNIE AUFFENBERG
BERNHARD KITTEL

No. 162

Universität Bremen	• University of Bremen
Jacobs Universität Bremen	• Jacobs University Bremen
Universität Oldenburg	• University of Oldenburg

Staatlichkeit im Wandel	• Transformations of the State
Sonderforschungsbereich 597	• Collaborative Research Center 597

*Kamil Marcinkiewicz
Jennie Auffenberg
Bernhard Kittel*

***Politikpositionen im Reformprozess des öffentlichen Dienstes
Zur Übertragbarkeit der quantitativen Textanalyse***

TranState Working Papers

No. 162

Sfb597 „Staatlichkeit im Wandel“ – „Transformations of the State“

Bremen, 2012

[ISSN 1861-1176]

Kamil Marcinkiewicz, Jennie Auffenberg, Bernhard Kittel

Politikpositionen im Reformprozess des öffentlichen Dienstes – Zur Übertragbarkeit
der quantitativen Textanalyse

(TranState Working Papers, 162)

Bremen: Sfb 597 „Staatlichkeit im Wandel“, 2012

ISSN 1861-1176

Universität Bremen

Sonderforschungsbereich 597 / Collaborative Research Center 597

Staatlichkeit im Wandel / Transformations of the State

Postfach 33 04 40

D - 28334 Bremen

Tel.: +49 421 218-8720

Fax: +49 421 218-8721

Homepage: <http://www.staatlichkeit.uni-bremen.de>

Diese Arbeit ist im Sonderforschungsbereich 597 „Staatlichkeit im Wandel“, Bremen, entstanden und wurde auf dessen Veranlassung unter Verwendung der ihm von der Deutschen Forschungsgemeinschaft zur Verfügung gestellten Mittel veröffentlicht.

Deutsche
Forschungsgemeinschaft

DFG

Politikpositionen im Reformprozess des öffentlichen Dienstes Zur Übertragbarkeit der quantitativen Textanalyse

ZUSAMMENFASSUNG

Der Aufsatz befasst sich mit der Übertragung eines Verfahrens der Quantitativen Inhaltsanalyse, das für die Untersuchung ideologischer Texte entwickelt wurde und hier auf die Diskussion seitens der Arbeitgeber- und Arbeitnehmervertretung zum Reformprozess des öffentlichen Dienstes in Deutschland angewendet wird. Es werden theoretische Argumente vorgestellt, die für die Anwendbarkeit des räumlichen Modells der Politik auf die Bestimmung der Positionen der Interessenvertretungen sprechen. Des Weiteren wird eine empirische Untersuchung am Beispiel des Polizeisektors auf Länderebene durchgeführt. Die Analyse der aus der Mitgliederzeitschrift der Gewerkschaft der Polizei und der Publikationen des Innenministeriums des Landes Brandenburg entnommenen Kommentare ermöglicht die Bestimmung von Politikpositionen im Zeitverlauf, die mit den Erkenntnissen anderer Studien übereinstimmen.

INHALT

EINLEITUNG.....	1
METHODE	3
Das räumliche Modell	3
„Wordscores“ als Beispiel eines Skalierungsverfahrens	5
Die Übertragung auf Reformen in der öffentlichen Verwaltung	7
ANALYSE	10
Auswahl der Daten	10
Auswahl von Referenzpunkten	14
Empirische Ergebnisse	16
„Word Scores“ und räumliche Positionen von Worttypen	16
„Text Scores“ und zeitliche Veränderung von politischen Positionen	21
ZUR ÜBERTRAGBARKEIT DES VERFAHRENS UND AUSBLICK	24
ANHANG	26
LITERATUR	30
BIOGRAPHISCHE ANMERKUNG.....	33

ABBILDUNGSVERZEICHNIS

Abbildung 1: Verteilung neutraler Worttypen im Datensatz	26
Abbildung 2: Zusammenhang zwischen Log-Häufigkeiten und „ideologischen“ Gewichten einzelner Worttypen.....	26
Abbildung 3: Zusammenhang zwischen Log-Häufigkeiten und „ideologischen“ Gewichten einzelner Worttypen (Auswahl extrem gewichteter und häufig verwendeter Worttypen)	27
Abbildung 4: Verteilung der für das IMB charakteristischen Worttypen	27
Abbildung 5: Verteilung der für die GdP charakteristischen Worttypen	28
Abbildung 6: Positionen der einzelnen Dokumente nach der MV-Transformation.....	28
Abbildung 7: Lowess-Smoother mit MV-transformierten Werten	29
Abbildung 8: Positionen der jahresweise aggregierten Dokumente (MV- Transformation)	29

Politikpositionen im Reformprozess des öffentlichen Dienstes Zur Übertragbarkeit der quantitativen Textanalyse¹

EINLEITUNG

Der rapide Fortschritt auf dem Gebiet der Computertechnologie in den letzten Dekaden blieb nicht ohne Folgen für die Politikwissenschaft. Neue rechnergestützte Verfahren automatisierten die Datenauswertung und verbesserten die Reliabilität von Analyseergebnissen. Die Folgen sind primär dort sichtbar, wo mit traditionellen statistischen Daten gearbeitet wird. Durch neue Programme können komplexe Analysen schneller und mit geringerem Personal- und Zeitaufwand durchgeführt werden.

Noch interessanter sind aber die Entwicklungen in Bereichen, die bisher als schwer zugänglich für automatische Verfahren galten, wie die Analyse politischer Texte. Die Idee, Wörter wie statistische Daten zu behandeln und entsprechend auf standardisierte Weise zu untersuchen, eröffnete auch für die Politikwissenschaft die Möglichkeit der quantitativen Inhaltsanalyse, die bisher insbesondere im Bereich der Parteienforschung Anwendung gefunden hat. So wurden innovative Ansätze wie „Wordscores“ und „Wordfish“ erstmals im Kontext des räumlichen Modells des Parteienwettbewerbs von Downs (1957) und Black (1958) durch Laver, Benoit und Garry (2003) sowie Slapin und Proksch (2008) angewendet. Ein großer Vorteil liegt in der Möglichkeit, latente Zusammenhänge sichtbar zu machen, d.h. Politikpositionen in ein räumliches Verhältnis zueinander zu setzen, die aus den Texten nicht direkt herauszulesen sind.

Das „Wordscores“-Verfahren lässt sich jedoch auch zur Darstellung politischer Positionen von Interessengruppen nutzen. Für die Positionen der hier untersuchten Interessengruppen – Arbeitgeber- und Arbeitnehmervertretungen – liegen allerdings keine klar umrissenen und aufgrund identischer Zielsetzung gut miteinander vergleichbare Textkorpora analog zu Parteiprogrammen vor. Daher wird ähnlich wie bei Klemmensen, Hobolt und Hansen (2007) auch hier eine Ausweitung der Methode auf eine bisher nicht im „Wordscores“-Verfahren verwendete Textart vorgenommen. Zudem wird auch die Übertragbarkeit des „Wordscores“-Verfahrens auf einen neuen Kontext, den des Aushandlungsprozesses widerstreitender Interessengruppen im Zeitverlauf überprüft. Eine besondere Herausforderung wird die neue Textart mit sich bringen, da sie die Kriterien Laver, Benoit und Garrys (2003) für die zu analysierenden Texte, d.h. dass sie in ihrem Umfang einheitlich und in ihren Inhalten kontrovers und „ideologisch“ aufgeladen sein müssen, im strengen Sinne nur beschränkt erfüllen.

¹ Wir danken Lisa Guta für ihre Unterstützung bei der Dateneingabe und Kodierung.

Der Begriff „ideologisch“ wurde in vorangehenden Studien zu Parteipositionen genutzt. Da diese Untersuchung darauf Bezug nimmt, wird der Begriff auch hier verwendet um die diskriminierende Wirkung bestimmter Worttypen aufgrund ihrer „Ladung“ (genauer hierzu unter METHODE) zu verdeutlichen. Dies geschieht, obwohl im Falle der Reformen im Rahmen des Neuen Steuerungsmodells (NSM) die unterschiedlichen Positionierungen der Akteure nicht auf unterschiedliche Ideologien im Sinne eines kohärenten politischen Ideengebäudes zurückgeführt werden können, sondern vielmehr auf unterschiedliche Politikpositionen in diesem spezifischen Kontext.

Inhaltlich zielt die Analyse auf die Erfassung der ideologischen Positionen der Arbeitgeber und Arbeitnehmer im öffentlichen Dienst zu zentralen Fragen von Reformen im öffentlichen Dienst, die sich auf die Beschäftigungsverhältnisse öffentlich Bediensteter auswirken. Finden sich in der Literatur umfangreiche Studien zu Umsetzung und Auswirkung der Verwaltungsreformen des öffentlichen Dienstes (Bogumil et al. 2008) sowie zu den in diesem Kontext angewendeten Formen der Beteiligung der Arbeitnehmerseite (Kißler, Graf und Wiechmann 2000) insbesondere auf kommunaler Ebene, so fehlt eine systematische Darstellung des Aushandlungsprozesses und der Positionierung der beteiligten Akteure im Reformprozess insb. zur Veränderung des Beschäftigungsregimes vom Idealtypus des Staatsdieners zum Idealtypus des Dienstleisters. Im vorliegenden Paper soll ein erster Schritt in diese Richtung unternommen werden. Mithilfe der „Wordscores“-Analyse² soll exemplarisch auf Landesebene die Positionierung der Gewerkschaft der Polizei sowie des Innenministeriums Brandenburg zu den Reformen des öffentlichen Beschäftigungsregimes im Zeitraum von 2002 bis 2010 nachgezeichnet werden. Dies soll Aufschluss über die Hintergründe des Aushandlungsprozesses von Verwaltungsreformen im öffentlichen Dienst der Polizei geben. Es wird zu zeigen sein, dass die hier verwendeten Texte trotz ihrer relativ schwachen Kontroversität und ihrer unterschiedlichen Länge dennoch mithilfe des „Wordscores“-Verfahrens Aufschluss über einen Trend in den Positionen der Arbeitgeber- und Arbeitnehmervertretungen geben können.

Dazu stellen wir kurz unsere theoretische Grundlage – das räumliche Modell – sowie das von uns genutzte Skalierungsverfahren der quantitativen Inhaltsanalyse, „Wordscores“ vor und zeigen, wie diese auf den Reformprozess im öffentlichen Dienst und die

² Die „Wordfish“-Analyse wird zunächst vernachlässigt, da es sich im untersuchten Fall um die Veränderung der Positionen von lediglich zwei Akteuren im Zeitverlauf handelt, was zufriedenstellend mit Hilfe von „Wordscores“ beschrieben werden kann. Überdies ist die Kontrolle über den Auswertungsprozess aufgrund der externen Setzung der Referenzpunkte hier höher, was insbesondere im Rahmen einer explorativen Analyse und bei der Untersuchung eines neuen Gebiets als ein Vorteil gesehen werden kann. Überdies können auf diese Weise die Texte aus dem ersten Jahr in unserem Datensatz als Referenz ausgewählt werden und die Entwicklung der Positionen im Zeitverlauf im Verhältnis zu diesen „Nullpunkten“ leichter interpretiert werden.

Aushandlungsprozesse der Arbeitgeber- und Arbeitnehmervertretungen anwendbar sind (*Methode*). In einem weiteren Schritt beschreiben wir die durchgeführte Analyse, beginnend mit der Auswahl der Daten und der für die „Wordscores“-Analyse essentiellen Setzung der Referenzpunkte, auf die anschließend die Darstellung und Interpretation der Ergebnisse folgt (*Analyse*). Zuletzt wird eine kritische Reflektion der Analyse vorgenommen und auf praktische wie theoretische Probleme eingegangen (*Zur Übertragbarkeit des Verfahrens und Ausblick*).

METHODE

Das räumliche Modell

Die Wurzeln des räumlichen Modells der Politik sind in den Arbeiten von Hotelling (1929) und Smithies (1941) zu finden, die sich primär mit der Standortwahl durch die Unternehmen beschäftigten (Hinich und Munger 1997: 21). Ihre Untersuchungen lieferten eine Antwort auf die Frage, warum konkurrierende Firmen häufig nebeneinander ansässig sind. Zugleich stellte schon Hotelling (1929) fest, dass sich die Logik der Standortwahl auf die Parteienkonkurrenz im US-amerikanischen Kontext übertragen lässt, wenn anstatt des geographischen Raumes, über einen ideologischen Raum gesprochen wird (Hotelling 1929: 55). Der Erforschung dieser Frage widmeten sich Downs (1957) und Black (1958), deren Werke die Grundlagen des räumlichen Modells der Politik legten (Hinich und Munger 1997: 21; Shepsle 2010: 111 f.). Das räumliche Modell wird in den erwähnten Werken als ein theoretischer Ansatz verstanden, der sich mit der Erfassung von Positionen politischer Akteure im ein- beziehungsweise multidimensionalen Raum beschäftigt.

Ein Mehrwert des räumlichen Modells besteht in seiner Übersichtlichkeit und theoretischen Exaktheit. Gleichzeitig war es von Anfang an mit dem Problem seiner empirischen Überprüfbarkeit konfrontiert. Wenn es ein Parteienspektrum ganzheitlich beschreiben soll, kann es sich nicht nur mit der Diskussion über die Höhe von Staatsausgaben beschäftigen (Black 1958: 3), die auf einer metrischen Skala problemlos abgebildet werden können. Diese einfache Übertragung funktioniert für viele andere politische Debatten nicht, da die Politikpositionen in der Regel auf Nominalskalenniveau modelliert werden müssten. Diese Positionen können jedoch oft durch Ideologien beschrieben werden, die in einem ein- bzw. multidimensionalen Koordinatensystem abbildbar sind.³ Ideologien sind jedoch hochabstrakte Phänomene, die sich nicht einfach messen lassen. Als eine latente Variable können sie lediglich indirekt anhand von Beobachtungen der

³ Zur Kritik des Links-Rechts-Ansatzes vgl. Albright (2010).

Handlungsweisen von Interessenvertretungen und ihrer Mitglieder geschätzt werden (Proksch und Slapin 2009: 324).

Eine Möglichkeit der Erfassung dieser latenten Variable bietet die Auswertung anhand eines eigenen Kategoriensystems (Coppedge 1997). Diese Methode weist jedoch Schwachstellen im Bereich der Objektivität und Replizierbarkeit auf.

Eine häufiger genutzte und vermutlich validere Option sind Experteninterviews (Benoit und Laver 2006; Castles und Mair 1984; Huber und Inglehart 1995; Wiesehomeier und Benoit 2009). Die Meinungen der Expertinnen und Experten sind jedoch ähnlich wie auch Befragungsergebnisse sehr stark durch frühere Erfahrungen beeinflusst (Gabel und Huber 2000). Die Ergebnisse solcher Untersuchungen spiegeln nicht wider, wie die Akteure sich selbst sehen und sich nach außen zeigen wollen, sondern wie sie durch externe Beobachter wahrgenommen werden.

Eine Alternative zur Bestimmung von Positionen in einem politischen Raum ist die Untersuchung politischer Texte. Der Nachteil dieses Ansatzes ist, dass diese nur das wiedergeben, was die politischen Akteure über sich selbst sagen. Auf der anderen Seite sind diese Inhalte beständig und können wiederholt mit unterschiedlichen Methoden untersucht werden, was für das Wissen der Experten nur begrenzt zutrifft. Anstatt sich auf die Interpretation des Vermittlers zu verlassen, wird hier direkt auf die Originalquelle zurückgegriffen. Wenn es gelingt, die Extraktion der politischen Positionen aus den Texten zu standardisieren, kann dies ein robusteres Analyseinstrument als die Auswertung anhand eines Kategoriensystems oder mithilfe eines Experteninterviews sein. Es kann argumentiert werden, dass durch die erworbene Replizierbarkeit und Reliabilitätsverbesserung Verluste in der externen Validität kompensiert werden.⁴ Bekannt geworden ist die quantitative Textanalyse durch die Analyse von Parteiprogrammen. Diese eignen sich in besonderer Weise, da sie in der Regel im Zuge interner Diskussionen entstehen, in denen unterschiedliche Parteiflügel einen Kompromiss erreichen (Ezrow et al. 2010: 280). Aus diesem Grund können diese Dokumente als repräsentativ für eine gegebene politische Gruppierung gelten.

Die Anfänge der systematischen Untersuchung von Parteiprogrammen liegen in den 1970er Jahren. Zwischen 1979 und 1989 kodierte die Manifesto Research Group Parteiprogramme aus 20 europäischen Staaten mit Hilfe eines Kategorienschemas, wobei einzelne Kategorien mit dem linken beziehungsweise rechten politischen Spektrum verknüpft wurden (Budge et al. 2001: 95). Das Projekt wurde nach 1989 unter dem Namen Comparative Manifesto Project auf 52 Staaten ausgeweitet. Die Kritik an der Reliabilität der manuell durchgeführten Kodierung führte zur Entwicklung computerge-

⁴ Es kann jedoch angemerkt werden, dass die externe Validität von Experteninterviews nicht notwendigerweise höher ist als in der Untersuchung von Parteiprogrammen, da fraglich bleibt, ob die Motivation der Handelnden von Experten erfasst werden kann.

stützter Verfahren. Im Jahre 2003 wurde „Wordscores“ präsentiert (Laver, Benoit und Garry 2003) und den Forschern in der Form eines STATA-Pakets zur Verfügung gestellt. Als eine weitere Alternative schlugen Slapin und Proksch fünf Jahre später „Wordfish“ vor (Slapin und Proksch 2008). Beide Methoden sind in Will Lowes R-Paket „Austin“ zugänglich, der zusätzliche Modifikationen in „Wordfish“ einführte, die die Schätzung der Standardfehler beschleunigen und qualitativ verbessern (Lowe 2010). Die in diesem Aufsatz präsentierte Analyse wurde mit Hilfe von „Austin“ durchgeführt.

„Wordscores“ als Beispiel eines Skalierungsverfahrens

Durch die Automatisierung der Datenauswertung mit Hilfe computergestützter Verfahren wurde das Problem der Inter- und Intracoder-Reliabilität in diesem Analyseschritt gelöst. Sowohl „Wordscores“ als auch „Wordfish“ behandeln Wörter als Daten (Benoit, Laver und Mikhaylov 2009). Die aufwändige Identifikation der Analyseeinheiten, wie zum Beispiel „Quasi-Sentences“ im Comparative Manifesto Project (Klingemann et al. 2006), entfällt dadurch. Weder „Wordscores“ noch „Wordfish“ erfordern ein Kategoriensystem oder ein externes Wörterbuch. So fungieren in der „Wordscores“-Analyse die Referenztexte als eine Art Wörterbuch. Dem Forscher wird aber weiterhin die Entscheidung über die Auswahl des Text-Korpus überlassen. Während dies im Falle von Parteiprogrammen selten kontrovers ist, muss bei der Anwendung des Verfahrens auf andere Gebiete eine klare Begründung für die Auswahl der für die Analyse bestimmten Dokumente geliefert werden. „Wordscores“ erfordert zusätzlich eine externe Angabe von mindestens zwei Referenztexten, mit denen die zu analysierenden Quellen („virgin texts“) verglichen werden (Laver, Benoit und Garry 2003).

„Wordscores“ bricht mit der Wahrnehmung der Texte als Diskurse, die als Ganzes interpretiert werden müssen und betrachtet sie stattdessen als Sammlung von Worten, die als Daten betrachtet werden können (Laver, Benoit und Garry 2003: 312). Die Worte beinhalten Informationen über die räumlichen Positionen der Texte, in denen sie sich befinden. „Wordscores“ leitet diese Informationen durch die Überprüfung der relativen Häufigkeiten ab, mit denen einzelne Begriffe in den Dokumenten auftauchen (ebd.). Die Vorteile von „Wordscores“ bestehen in der mathematischen Einfachheit des Verfahrens, das keine expliziten Verteilungsannahmen benötigt (Lowe 2008: 356). Zugleich aber ermöglicht „Wordscores“ die Berechnung von Konfidenzintervallen für bestimmte Policy-Positionen. In der praktischen Anwendung fängt die Analyse mit der Auswahl von Referenztexten an, die extern vorgenommen werden muss. Im zweiten Schritt werden die Gewichtungen von einzelnen Begriffen bestimmt, welche sich anhand der relativen Häufigkeiten des Auftauchens in einzelnen Referenztexten ermitteln lassen. Im dritten Schritt erfolgt die Bestimmung der Policy-Positionen der untersuchten Dokumente. Die

Herausforderungen, die sich in den oben genannten Phasen des Verfahrens ergeben, werden in den folgenden Abschnitten des Aufsatzes besprochen.

„Wordscores“ wurde für die Analyse von Parteiprogrammen entwickelt und erfolgreich eingesetzt. Im Vergleich mit den Schätzungen von Experten zeigt die im automatisierten Verfahren durchgeführte Positionierung einen höheren Grad der Übereinstimmung als die manuelle Vercodung. Die Werte, die Laver, Benoit und Garry (2003: 320) für die britischen Parteien ermittelten, liegen nicht nur nahe an den Experteneinschätzungen, sondern erfassen darüber hinaus politische Phänomene, wie den liberaleren Charakter des Wirtschaftsprogramms der Labour Party im Vergleich mit Aussagen der Liberaldemokraten. Die Erfinder der Methode kommen zu dem Schluss, dass „Wordscores“ effizienter als die manuelle Kodierung ist und dies, ohne sich inhaltlich mit dem im analysierten Dokument geführten Diskurs zu befassen (Laver, Benoit und Garry 2003: 321).

Neben der Anwendung auf die Analyse von Parteiprogrammen prüften schon die Erfinder des Verfahrens die Möglichkeit seiner Übertragbarkeit auf andere Arten politischer Texte. So analysierten Laver und Benoit (2002) mit Hilfe von „Wordscores“ die Haushaltdebatte im irischen Parlament. Anstatt die Positionen von Parteien auszuwerten, bestimmen sie die politischen Profile der einzelnen Abgeordneten. Hier stellte sich das Problem, dass bei der Analyse der Parlamentsdebatten nicht auf die Logik, die für die Bestimmung von Referenztexten für Parteiprogramme angewendet wird, zurückgegriffen werden kann. Im Falle der Analyse von Parteipositionen werden Programme aus den vergangenen Wahlen als Bezugspunkte genommen, denen auf der Grundlage von Experteninterviews Gewichte zugewiesen werden (Coffe und Plassa 2010). Während analog Positionen von Abgeordneten aus Studien früherer Haushaltsdebatten als Referenzpunkte genutzt werden könnten, schlagen Laver und Benoit als Referenzpunkte Reden von Parteiführern vor, welche dann als beispielhaft für gegensätzliche Positionen erklärt und mit jeweils -1 und 1 gewichtet werden.

Mit einem ähnlichen Dilemma setzen sich Klemmensen, Hobolt und Hansen (2007) auseinander, die die politischen Positionen der dänischen Regierungen zwischen 1953 und 2006 untersuchen. Ihr Datensatz besteht aus den Erklärungen der Regierungschefs, die in einer Zeitreihe erfasst werden. Als Referenzpunkte werden Texte aus unterschiedlichen Jahren genommen, die als eindeutig links beziehungsweise rechts gelten. Versuche mit unterschiedlichen Bezugspunkten bestätigen die Existenz einer langfristigen Rechtsverschiebung in der dänischen Politik.

Klüver (2009) untersucht in ihrer Arbeit die Anwendbarkeit unterschiedlicher Formen quantitativer Textanalyse auf die Positionierung von Interessengruppen im Kontext eines EU-Gesetzgebungsprozesses. Gegenstand ihrer Untersuchung ist die Debatte über die geplante Initiative der Kommission zum Abbau der von Kraftfahrzeugen verursach-

ten CO₂-Emissionen. In diesem Zusammenhang kann die Dimension des Konfliktes relativ einfach identifiziert werden. Die Interessengruppen können den Vorschlag der Kommission entweder unterstützen oder ihn ablehnen. Dies erlaubt die Anwendung von „Wordfish“. Gleichzeitig ist aber die Nutzung von „Wordscores“ problematisch, denn es fehlen klare Referenzpunkte.

In der vorliegenden Studie wird sowohl an die Erfahrungen von Laver und Benoit (2002) als auch von Klemmensen, Hobolt und Hansen (2007) angeknüpft. Dabei werden die Einschränkungen, auf die Klüver (2009) aufmerksam macht, berücksichtigt. Wir analysieren Dokumente der Gewerkschaft und der Arbeitgeber, die mit zwei Arten der Wahrnehmung des öffentlichen Dienstes assoziiert werden – dem Idealtypus des Staatsdieners und dem Idealtypus des Dienstleisters. Diese zwei Idealtypen bilden die extremen Pole des untersuchten Policy-Raums. Die Untersuchung soll die Veränderung der Positionen der zwei Akteure im Zeitverlauf erfassen, weshalb sich Dokumente beider Akteure zu Anfang der Zeitreihe am besten als Referenztexte eignen.

Die Übertragung auf Reformen in der öffentlichen Verwaltung

Wie bereits erwähnt hat Klüver (2009) gezeigt, dass das räumliche Modell zur Darstellung des Einflusses von Interessengruppen angewendet werden kann. Sie hat dafür – allerdings mithilfe von „Wordfish“ – die Präferenzen der Interessengruppen und deren Nähe zum Policy-Output verglichen (ebd: 546). Unser Erkenntnisinteresse ist ähnlich gelagert, da wir ebenfalls mit grundsätzlich opponierenden Interessenvertretungen – Gewerkschaften und öffentliche Arbeitgeberverbände – zu tun haben. Es geht jedoch nicht darum, den Einfluss dieser Akteure auf die Verwaltungsmodernisierungen zu messen, sondern deren Positionierung zu den Verwaltungsreformen im öffentlichen Dienst bzw. zu den Idealtypen des Staatsdieners und des Dienstleisters für den Zeitraum von 2002 bis 2010 in Relation zueinander darzustellen. Diese Positionierung können wir als eine latente Variable betrachten. Diese Positionen sind also aus den Publikationen nicht direkt ablesbar, vielmehr werden sie erst durch die „Wordscores“-Analyse in einem räumlichen Modell sichtbar. Der Verlauf der für die einzelnen Publikationen bzw. aggregierten Publikationsjahrgänge erhaltenen Punkte soll Aufschluss über die Rolle der beiden Akteure im Reformprozess des öffentlichen Beschäftigungsregimes geben. So wird erwartet, dass die entstehende Kurve Hinweise darauf geben kann, ob und gegebenenfalls wie sich die ursprüngliche Einstellung der beiden Akteure zu den Verwaltungsreformen und den Idealtypen der öffentlichen Beschäftigung verändert hat. Zum Beispiel lässt sich fragen, ob die Gewerkschaften zu bestimmten Momenten kooperationsbereiter waren oder ob Reformen in Kooperation oder gegen den Willen der Gewerkschaften durchgesetzt wurden.

Da sich bei der konkreten Umsetzung des Forschungsdesigns nur wenige Texte finden ließen, die die Reformen des Beschäftigungsregimes betrafen und der Kurvenverlauf aufgrund zu weniger Punkte nicht interpretierbar gewesen wäre, musste die anfänglich angestrebte enge Betrachtung des Beschäftigungsregimes ausgeweitet werden. In der weiteren Analyse konnte nur die räumliche Positionierung Pro (der Arbeitgeberseite unterstellt) bzw. Contra (der Gewerkschaftsseite unterstellt) die Verwaltungsmodernisierungen im Allgemeinen dargestellt werden (mehr hierzu im Abschnitt ANALYSE).

Für die „Wordscores“-Analyse ist nach Auswahl der Referenztexte eine Gewichtung der anfänglichen Position der Interessengruppen nötig. Dies kann problematisch sein, wenn es – wie im Beispiel der Analyse der europäischen CO₂-Gesetzgebung - an Experteneinschätzungen fehlt (Klüver 2009: 547). Auch wenn für die Positionierung der Gewerkschaften und Arbeitgeberverbände diese Möglichkeiten ebenfalls nicht gegeben sind, ist dies für die hier beschriebene Analyse unerheblich, da sich die Untersuchung auf nur zwei Akteure konzentriert. Die Annahme, dass diese zu Beginn unterschiedliche Positionen vertreten haben, genügt für die dichotome Setzung der Referenzpunkte.

Allerdings muss diese Annahme begründet werden. Vor dem Hintergrund, dass die Verwaltungsmodernisierungen Ende der 80er Jahre in vielen Kommunen von den Gewerkschaften selbst angestoßen wurden (Semmler und Weber 2005: 75; Sperling 1998: 329) ist dies zunächst fraglich. So waren Ende der 80er und Anfang der 90er Jahre die öffentlichen Beschäftigten auf kommunaler Ebene selbst und insbesondere die Führungskräfte unter ihnen, aufgrund interner Probleme mit den Verhältnissen in der Verwaltung unzufrieden und eine Modernisierung der Verwaltung wurde als Alternative zu Privatisierungen und „neoliberalem Minimalstaat“ gesehen (Semmler und Weber 2005: 75). Die Qualität der öffentlichen Dienstleistungen sollte zur Sicherung der Arbeitsplätze führen und es wurde angenommen, dass diese auch von der Qualität der Arbeitsbedingungen abhängt (Schneider 2011: 359 f.). So hat die Gewerkschaft Öffentliche Dienste, Transport und Verkehr (ÖTV) schon 1988 ihre Reforminitiative „Zukunft durch öffentliche Dienste (ZöD)“ ins Leben gerufen, womit sie das Problembewusstsein für eine Reformnotwendigkeit und ihre Reformbereitschaft signalisiert hat (Frieß et al. 2000: 135). Der dahinter liegende Grundgedanke war, dass sich Verwaltungs- und Arbeitsqualität gemeinsam optimieren lassen (ebd.). Dass die Gewerkschaften kooperations- und reformbereit waren, belegen auch zahlreiche Dienstvereinbarungen zur Verwaltungsmodernisierung, die seit Mitte der 90er Jahre auf kommunaler Ebene zwischen Personalräten, Gewerkschaften und der politischen oder administrativen Führung ausgehandelt wurden (Sperling 1998: 334 f.). Offen ist jedoch, insbesondere im Hinblick auf die Alternative der Privatisierung, wie erfolgreich sie bei der Durchsetzung ihrer Strategie waren, und wie die Gewerkschaft den Zielkonflikt zwischen Strategie- und

Verhandlungsfähigkeit und der Wahrung der Interessen ihrer Klientel im öffentlichen Dienst diskursiv in der Öffentlichkeit positioniert hat.

Infolgedessen und aufgrund der Erkenntnis, dass eine Verwaltungsmodernisierung ohne die Unterstützung der Arbeitnehmer nicht umsetzbar ist, wurden diese in Form der Beschäftigten, der Gewerkschaften und insbesondere des Personalrats in die Gestaltung des Reformprozesses auf kommunaler Ebene einbezogen (Bogumil et al. 2008: 113 ff.; Schneider 2011: 359; Sperling 1998: 324). Waren dadurch zwar vielfältige und z.T. auch umfangreiche Partizipationsmöglichkeiten gegeben, so haben jedoch die beteiligten Beschäftigten im Laufe des Reformprozesses feststellen müssen, dass sie zwar am Prozess beteiligt waren, ihr Einfluss auf die Gestaltung der Verwaltungsreformen jedoch nur gering war (Bogumil et al. 2008: 109 ff.; Frieß et al. 2000: 136 ff.). Wenn dies nach Bogumil et al. auch nicht zu einer „Aussteigermentalität“ (Kißler, Graf und Wiechmann 2000: 70) geführt hat, so ist dennoch festzustellen, dass sich dieser Tatbestand in Demotivation und Partizipationsverweigerung sowie der Ablehnung neuer Reformanläufe ausdrückt (Bogumil et al. 2008: 113 ff., 305).

Deutlich wird, dass sich die beiden Akteure zwar anfänglich einig über die Notwendigkeit von Verwaltungsreformen waren und eine hohe Kooperationsbereitschaft auf beiden Seiten bestand. Daher kann nicht davon ausgegangen werden, dass ihre Positionen auf der Grundlage von Referenztexten aus dem Jahre 1990 in der Analyse dichotom dargestellt werden können. Wie jedoch festgestellt wurde, hat sich die Partizipationsbereitschaft bereits um die Jahrtausendwende seitens der Arbeitnehmer gewandelt, so dass angenommen werden kann, dass sich die Positionen der Akteure in unseren Referenzpunkten im Jahre 2002 unterscheiden.

Können die Referenzpunkte in Bezug auf die kommunale Ebene für das Jahr 2002 also begründet unterschiedlich gesetzt werden, so fällt die Argumentation für die Landesebene – hier dem Polizeisektor – schwerer, da entsprechende Literatur nicht in der gleichen Fülle wie für die kommunale Ebene vorhanden ist.⁵ Dennoch lässt sich ergänzen, dass der Reformprozess in den Polizeien der Länder erst fünf Jahre später als in den Kommunalverwaltungen, ab Mitte der 90er Jahre begann. Auch hier war das Neue Steuerungsmodell zunächst eine Forderung der Mitarbeiterinnen und Mitarbeiter gegenüber dem Management (Barthel 2008: 377 ff.) bzw. eine Reaktion desselben auf die Unzufriedenheit und veränderten Erwartungen der Beschäftigten (Schmidt 2005: 38 ff.). Doch nach der anfänglichen „Konzept euphorie“ stellte sich auch hier etwa im Jahr 2002/2003 Ernüchterung und Veralltäglichung ein. Dies war unter anderem auf die bürokratisch-autoritäre Anordnung der Reformen sowie auf Unsicherheiten und Ängste der betroffenen Beschäftigten zurückzuführen. Deutlich wurde, dass der maßgebliche

⁵ Zur Problematik der mangelnden wissenschaftlichen Aufarbeitung des Reformprozesses in den Polizeien der Länder siehe (Barthel 2008: 380 ff.).

Akteur im Reformprozess der höhere Dienst und das mittlere Polizeimanagement waren, d.h. der überwiegende Teil der Beschäftigten nicht einbezogen wurde (Barthel 2008: 377 ff.). Diese Feststellungen lassen noch keinen direkten Rückschluss auf die Positionen der Akteure zu. Dennoch zeigen sich Parallelen zum Reformprozess in der Kommunalverwaltung, so dass vorerst angenommen wird, dass auch für den Polizeisektor zu Anfang der vorliegenden Analyse im Jahr 2002 von unterschiedlichen Positionen ausgegangen werden kann.

Dieses Vorgehen ähnelt also dem Vorgehen Laver und Benoits (2002), die in ihrer Analyse der irischen Parlamentsdebatten von einer grundsätzlich unterschiedlichen Positionierung von Opposition und Regierung ausgingen und als Referenzpunkte auf der einen Seite die Rede des Regierungschefs und auf der anderen Seite die Reden der Anführer der beiden größten Oppositionsparteien auswählten. Innerhalb dieser Spannweite sollten die Reden der übrigen Parlamentsmitglieder eingeordnet werden.

ANALYSE

Auswahl der Daten

Für die vorliegende Analyse kann nicht auf regelmäßig erscheinende Programme von Gewerkschaften und Arbeitgeberverbänden zurückgegriffen werden, so wie es beispielsweise in der Analyse von Parteipositionen möglich ist. Gewerkschaften und Arbeitgebervertretungen konkurrieren zudem grundsätzlich nicht um die gleiche Zielgruppe, das heißt die gleichen Mitglieder. Doch im Bereich der Verwaltungsreformen ist auch der öffentliche Arbeitgeber auf die Kooperation der Beschäftigten angewiesen (Bogumil et al. 2008: 113 ff.), so dass davon ausgegangen werden kann, dass hier auch die Arbeitgeberseite um die Zustimmung der Beschäftigten wirbt. So findet sich das Thema der Verwaltungsmodernisierung nicht nur in den Mitgliederpublikationen der Gewerkschaften wieder, sondern ebenso in den von Arbeitgeberseite herausgegebenen Mitarbeiterzeitschriften. Dies begründet auch die Auswahl eben dieser Medien für unsere Analyse. Im Detail handelt es sich hier um die monatlich erscheinende Zeitschrift der Gewerkschaft der Polizei (GdP) „Deutsche Polizei“ als Mitgliederzeitschrift der größten deutschen Polizeigewerkschaft, sowie die ca. dreimal jährlich erscheinende Zeitschrift „info 110“ des Innenministeriums Brandenburg (IMB), exemplarisch für die öffentlichen Arbeitgeber der auf Länderebene organisierten Polizei.⁶

⁶ Brandenburg wurde aus zwei Gründen gewählt. Zum einen gehört Brandenburg neben Bremen und Baden-Württemberg zu den im Teilprojekt D6 des SFB 597 näher untersuchten Bundesländern. Zum anderen existierten nur für dieses Bundesland vollständig archivierte und uns zugängliche Publikationen.

Bei der Auswahl der „virgin“-Texte für die „Wordscores“-Analyse ist die gleichzeitige Sicherung der formalen Vergleichbarkeit sowie einer gewissen inhaltlichen Heterogenität von Bedeutung. Nur Dokumente, die sich in Form und Inhalt ähneln, jedoch unterschiedliche Meinungen präsentieren, eignen sich für das hier verwendete Verfahren. Wenn ein Minimum an Vergleichbarkeit nicht gegeben ist, kann die grundlegende Policy-Dimension nicht eindeutig ermittelt werden. Die Auswertung nichtkontroverser Texte würde auf die Identifizierung formaler Unterschiede zwischen Texten hinauslaufen.

Innerhalb der ausgewählten Publikationen erschienen daher die Kommentare als formales Kriterium der Untersuchungseinheiten geeignet. Es wurde eine Auswahl der Textart nach den Kriterien des Kommentars getroffen, eine sogenannte „syntactical distinction“ (Krippendorff 2004: 103). Diese Textart ist zunächst eindeutig zu identifizieren und ihre Struktur und ihr Umfang bleiben für die jeweiligen Publikationen im Zeitverlauf konstant. Somit ist eine gewisse Vergleichbarkeit gegeben, die Ergebnisse werden nicht aufgrund äußerer Merkmale verfälscht. Des Weiteren sind Kommentare – ähnlich wie etwa die bei Klemmensen, Hobolt und Hansen (2007: 747) untersuchten Reden der dänischen Regierungschefs – meinungsbetont und in ihnen findet eine Bewertung des behandelten Gegenstands statt (Nowag 2005: 148 ff.). Genau diese Wertung ist grundsätzlich geeignet, die nötige Heterogenität zu erzeugen und ist für die vorliegende Analyse relevant. Nur durch sie lassen sich die Positionen der Akteure unterscheiden, die später im räumlichen Modell sichtbar werden sollen. Die reine Beschreibung der Sachlage durch Arbeitgeberverbände und Gewerkschaften unterscheidet sich weniger voneinander als die bewusste Kommentierung der Geschehnisse und wäre in diesem Fall nicht zielführend. Zudem besitzen die Kommentare auch deshalb eine besondere Relevanz, da sie fast ausschließlich unter dem Namen der Vorsitzenden oder anderer hochrangiger Vertreter der jeweiligen Organisation veröffentlicht werden und somit anzunehmen ist, dass sie für die jeweiligen Organisationen als Ganze sprechen.

Da sich diese Kommentare jedoch nicht ausschließlich mit den Reformen im öffentlichen Dienst beschäftigen, wurde auch innerhalb der Kommentare eine inhaltliche Auswahl durch drei in dem Projekt involvierte und geschulte Codierer vorgenommen. Die Auswahl der Stichprobe orientierte sich an einem hierfür erstellten Kategoriensystem, wodurch gleichzeitig eine Codierung vorgenommen wurde. Im Grunde handelt es sich hier also um eine Vollerhebung – zunächst werden alle Kommentare gesichtet, die im untersuchten Zeitraum in den zu analysierenden Publikationen erschienen. Daraufhin werden für die quantitative Analyse jedoch nur die Texte verwendet, die sich mit den für die Forschungsfrage relevanten Themen befassen – es wird ein „purposive or judgment sampling“ (Neuendorf 2002: 88) oder auch „systematic sampling“ (Krippendorff 2004: 115) vorgenommen.

Das Kategoriensystem stützt sich zum einen auf die Klassifizierung der auf das Beschäftigungsregime bezogenen Reformen nach Gottschall und Kittel (2007) sowie auf die Klassifizierung der Reformen des Neuen Steuerungsmodells allgemein nach Bogumil et al. (2008). Gottschall und Kittel beschreiben zwei gegensätzliche Idealtypen der öffentlich Beschäftigten, den Staatsdiener- und den Dienstleister-Typus. Diese werden durch bestimmte Charakteristika des Beschäftigungsregime hervorgebracht (Gottschall und Kittel 2007: 29; Hils und Streb 2010: 8 ff.). Texte, die in die Kategorien der Reformen des Beschäftigungsregimes fallen, stellen den Kern unseres Interesses dar. Die Auswahl wurde jedoch um die weitreichendere Kategorisierung nach Bogumil et al. ergänzt, da dieses System alle mit dem Neuen Steuerungsmodell in Zusammenhang stehenden Reformen erfasst und so deutlich mehr Texte in die Analyse aufgenommen werden konnten.⁷ Durch die Analyse aller Kommentare zu den Verwaltungsreformen verändert sich die Dimension unseres räumlichen Modells. Das Beschäftigungsregime mit den Extrempunkten des Staatsdiener- und des Dienstleistertypus kann nicht mehr dargestellt werden. Die Dimension bildet nun die Zustimmung bzw. Ablehnung der Verwaltungsmodernisierungen im Allgemeinen ab. Dennoch wird zunächst eine detaillierte Kategorisierung der Kommentare zu den Verwaltungsreformen vorgenommen.

Die Kategorisierung der Reformen, die sich auf das Beschäftigungsregime beziehen, gliedert sich in drei Bereiche: Die Regulierung der Beschäftigungsverhältnisse (1) umfasst die Rechtsform sowie den Status des Beschäftigungsverhältnisses als auch deren Zustandekommen, d.h. die tarifvertragliche Regulierung und die Regulierung der Arbeitskonflikte. Die Personalpolitik (2) umfasst zum einen die Spezifika des öffentlichen Dienstes im Bereich der Rekrutierung, der Zugangsvoraussetzungen und der Austauschfähigkeit des öffentlichen Dienstes mit der Privatwirtschaft. Des Weiteren werden in dieser Kategorie auch das Personalmanagement wie Beurteilung, Leistungsanreize und Leistungskontrolle sowie die Personalentwicklung, d.h. Weiterbildungs- und Aufstiegsmöglichkeiten erfasst. Mit der Vorbildfunktion des öffentlichen Beschäftigungsregimes (3) gelangen alle die Kommentare in die Analyse, die Entlohnung, soziale Sicherheit und Beschäftigungssicherheit ansprechen sowie die Integration benachteiligter gesellschaftlicher Gruppen wie Behinderte, Frauen, Migrantinnen und Migranten, Ältere, etc. (Gottschall und Kittel 2007: 29; Hils und Streb 2010: 8 ff.). So beschreibt beispielsweise der Kommentar „Altersmäßig fortgeschritten“ aus der „Deutschen Polizei“ vom April 2009 das Problem der Überalterung und der Nachwuchsgewinnung im Polizeidienst. Hier wird die Rekrutierung von Nachwuchskräften problematisiert und die

⁷ Eine Analyse allein mit Kommentaren, die das Beschäftigungsregime betreffen, brachte aufgrund der geringen Fallzahl nur schlecht interpretierbare, nicht signifikante Ergebnisse hervor.

Forderung gestellt, den Polizeidienst attraktiver zu gestalten. Dieser Kommentar würde der Kategorie der Personalpolitik zugeordnet werden.

Die beschriebene Kategorisierung des Beschäftigungsregimes wird um fünf Bereiche der NSM-Reformen im Allgemeinen ergänzt. Mit dem Verhältnis von Politik und Verwaltung (4) wird der Grad der Trennung dieser beiden Sphären beschrieben. Anstelle einer engen Zusammenarbeit treten Vereinbarungen durch politische Kontrakte und Kontrolle durch politisches Controlling. Zur Verwaltungsorganisation (5) zählen diejenigen Reformvorhaben, welche die Umstrukturierung der Verwaltung betreffen, wie u.a. die Dezentralisierung der Verantwortungsstrukturen und der damit einhergehenden Einführung einer zentralen Steuerungsunterstützung. In diese Kategorie fällt beispielsweise die Polizeistrukturereform aus 2010, die in der ersten Ausgabe der „info 110“ aus dem Jahr 2010 behandelt wird. In den Bereich des Verwaltungsmanagements (6) fallen alle Kommentare, die die Einführung betriebswirtschaftlicher Steuerungselemente behandeln, wie die Einführung des Produktkonzepts, der Budgetierung oder eines Berichtswesens. Dies fand sich beispielsweise im Kommentar „Investitionen in die Zukunft“ aus der ersten Ausgabe des Jahres 2007 wieder, in dem von der Evaluierung der Kriminalpolizei die Rede ist. Mit der Bürgerorientierung (7) wird die Sicht des Bürgers als Kunden und Nachfrager nach Dienstleistungen beschrieben. Damit verbunden ist die Einführung eines Qualitätsmanagements, Maßnahmen wie die Umstrukturierung zu One-Stop-Agencies oder die Steigerung der Geschwindigkeit von Vorgängen. Zuletzt wird durch dieses zweite Kategoriensystem auch die Einführung von Wettbewerbselementen (8) zum Anreiz von Innovationen und zur Verbesserung von Verwaltungsstrukturen und –prozessen berücksichtigt. Dies können beispielsweise Wettbewerbssurrogate sein, wie die interne Leistungsverrechnung oder das Kontraktmanagement, aber auch Leistungsvergleiche zwischen Kommunen bis hin zu Privatisierungen staatlicher Aufgaben (Bogumil et al. 2008: 23 ff.).

Beachtet werden muss auch in diesem Zusammenhang wieder, dass Bogumil et al. die Dimensionen des Neuen Steuerungsmodells anhand des öffentlichen Dienstes auf kommunaler Ebene entwickelt haben. Da vergleichbare Modelle für die Landes- und Bundesebene jedoch nicht vorhanden sind, wird das kommunale Modell analog auch auf diese Sachverhalte angewendet. Diese Übertragung ist möglich, da auch im Polizeisektor das Neue Steuerungsmodell der Kommunalen Gemeinschaftsstelle für Verwaltungsmanagement (KGSt) als Grundlage galt, auf die im Reformprozess immer wieder Bezug genommen wurde (Barthel 2008: 377 ff.; Frevel und Kuschewski 2009: 74).

Des Weiteren kann eingewendet werden, dass sich ein Kommentar, der zwar in eine Kategorie des Kategoriensystems fällt, nicht automatisch mit einer Reform in Zusammenhang mit dem Neuen Steuerungsmodell stehen muss. Ein solcher Kommentar kann aber dennoch in die Analyse aufgenommen werden, da einerseits das Neue Steuerungs-

modell als solches nie Gegenstand der Kommentare ist und andererseits die im Kategoriensystem beschriebenen Bereiche solche sind, in denen Modernisierungen stattgefunden haben. So stehen alle Kommentare, die in einen der Reformbereiche fallen zumindest indirekt mit den Verwaltungsmodernisierungen in Verbindung.

Die hier vorgenommenen Kategorisierungen sind weitaus detaillierter, als es für die „Wordscores“-Analyse nötig wäre – hier wäre eine Unterscheidung danach, ob sich die Texte mit Reformen im öffentlichen Dienst beschäftigen oder nicht, bereits ausreichend. Sie können aber zu einem späteren Zeitpunkt dazu genutzt werden, um eine Wörterbuch-basierte Analyse oder auch die „Wordscores“-Analyse in verschiedenen Dimensionen durchzuführen.

Insgesamt erschienen im untersuchten Zeitraum 108 Ausgaben der Zeitschrift „Deutsche Polizei“ und 25 Hefte von „Info 110“. Etwa die Hälfte (55 von 108) der Kommentare der Gewerkschaften wurden als relevant für die Analyse anerkannt. Die Coders stellten fest, dass auf der Arbeitgeberseite sich 10 von 25 Texten mit der Verwaltungsreform befassten.

Die Intercoder-Reliabilität wurde anhand der Berechnung von Cohen’s Kappa kontrolliert. Cohen’s Kappa vergleicht die tatsächliche Übereinstimmung zwischen den Auswahlen der Coder mit der Übereinstimmung, die per Zufall erreicht werden würde. Es wird in der Literatur als ein konservatives Gütemaß bezeichnet, das lediglich die nicht zufällig erreichte Übereinstimmung belohnt (Neuendorf 2002: 151). Für die hier relevante Unterscheidung der Texte danach, ob sie die Verwaltungsmodernisierungen im Allgemeinen betreffen, liegt Cohen’s Kappa im Falle der Kommentare der Zeitschrift „Deutsche Polizei“ bei 0,871. Die Codierer haben in 101 von 108 Fällen die gleiche Auswahl getroffen. Für die Zeitschrift des brandenburgischen Innenministeriums lag der Wert von Cohen’s Kappa bei 0,677. Von 25 Dokumenten wurden 21 übereinstimmend kategorisiert. Beide Werte werden in der einschlägigen Literatur als sehr akzeptabel angesehen, weil sie näher an einer perfekten Übereinstimmung (Kappa=1) als an einer zufälligen Übereinstimmung (Kappa=0) liegen (Neuendorf 2002: 151).

Auswahl von Referenzpunkten

Ähnlich wie Klemmensen, Hobolt und Hansen (2007) soll hier die zeitliche Entwicklung politischer Positionen untersucht werden. Es sollen zwei Zeitreihen erzeugt werden, eine für die Publikationen des brandenburgischen Innenministeriums, stellvertretend für die öffentliche Arbeitgeberseite, und eine für die Publikationen der GdP als Repräsentantin der Arbeitnehmerseite. Interessant in diesem Kontext ist die Veränderung der Positionen im Verhältnis zum Ausgangspunkt und zugleich zwischen den beiden untersuchten Akteuren. Aus diesem Grund werden als Referenztexte Dokumente aus dem ersten Zeitpunkt in der Zeitreihe benutzt. Dies ist in unserem Fall das Jahr

2002. Wegen des relativ knappen Umfangs der Texte werden alle inhaltlich relevanten Dokumente dieses Jahres für den Referenztext verwendet. Die inhaltliche Vorauswahl ergab, dass lediglich drei von zwölf Beiträgen aus der GdP-Zeitschrift und zwei von drei Kommentaren des Innenministeriums aus dem Referenzjahr die hier untersuchte Problematik tangierten. Die drei Texte der „Deutsche Polizei“, wie auch einer der beiden Beiträge der „info 110“ sprechen die Verwaltungsmodernisierungen zudem nur im Allgemeinen an. Die Analyse wird aufgrund der sonst zu geringen Anzahl an Texten, die sich für die Analyse eignen, zunächst die Positionierungen zum Neuen Steuerungsmodell im Allgemeinen untersuchen.

Die beiden Gruppen der Kommentare werden aggregiert (Bernauer und Bräuninger 2009). Auf diese Weise werden zwei Referenztexte erzeugt. Das Zusammenfügen der einzelnen Beiträge zu zwei Referenztexten ist durch drei Ziele motiviert. Erstens, sollen die Referenztexte so gut wie möglich das Vokabular des gesamten Text-Korpus widerspiegeln (Klüver 2009; Laver, Benoit und Garry 2003). Von daher ist die Nutzung von mehreren zusammengefassten Kommentaren empfohlen. Zweitens fügen wir die Texte auch aus pragmatischen Gründen zusammen. Im Falle von zwei Referenztexten ist die Logik von „Wordscores“ einfacher nachvollziehbar, was besonders wichtig für die Anwendung auf das neue Feld ist. Drittens muss bei der Zusammensetzung von einzelnen Kommentaren zu zwei Referenztexten keine Einschätzung der Position der einzelnen Texte vorgenommen werden. Dies lässt uns zum einen die Ungenauigkeit des Outputs minimieren und zum anderen ist dies einfacher mit dem Transformationsverfahren nach Martin und Vanberg (2008b) vereinbar, was unter „*Text Scores*“ und *zeitliche Veränderung von politischen Positionen* ausführlicher erklärt wird.

Die Gewichtung der Referenztexte wird extern bestimmt. Eine Skala mit dem Wert von 0 in der Mitte ist aus der Perspektive der Verfasser relativ einfach zu interpretieren und aus diesem Grund gut geeignet für die Analyse. Die Punkte -10 (Innenministerium Brandenburg) und 10 (Gewerkschaft der Polizei) wurden gewählt um die Logik des Dezimalsystems widerzuspiegeln und zugleich zu kleine und zu große Zahlen zu vermeiden. Die Auswahl von -1 und 1 würde dazu führen, dass die Positionen der einzelnen Texte in den meisten Fällen durch Bruchzahlen beschrieben würden. Bei der Skala zwischen -100 und 100, auf der anderen Seite, würden unnötig große Zahlen erzeugt, was bei zwei zu analysierenden Akteuren nicht notwendig ist.

Der Umfang des Referenztextes und die Häufigkeit, mit der ein bestimmter Worttypus im Dokument auftaucht, sind wichtig für die Qualität der Auswertung. Je mehr Worttypen der Referenztext umfasst, desto größer ist der Anteil des Text-Korpus, der in die Analyse einfließen kann („scorable words“). Ähnlich gestaltet sich das Problem der Wiederholungsrate. Da die Grundlage für die Berechnung der Gewichtung eines Worttyps die Häufigkeit seines Auftauchens im einen Referenztext im Verhältnis zum ande-

ren ist, steigt die Genauigkeit der Analyse mit der Anzahl der Wiederholungen von Worttypen. Dies kann durch das Gesetz der großen Zahlen begründet werden und betrifft „Wordscores“ in ähnlichem Maße wie auch andere quantitative Verfahren. Zusammenfassend gesagt: Je näher die Größe einer Stichprobe an der Grundgesamtheit liegt, desto realistischer bildet sie Verhältnisse ab, die für die Grundgesamtheit gelten (Bortz 2005: 93 f.).

Der Referenztext für die Arbeitgeberseite besteht aus 583 Worttypen und ist somit anderthalbmal umfangreicher als ein durchschnittlicher „virgin“-Text, entnommen aus der „Info 110“ (368 Worttypen). Ähnlich gestalten sich auch die Verhältnisse zwischen den ausgewerteten Texten und den für den Vergleich benutzten Dokumenten bei der GdP. Der aggregierte Referenztext enthält mehr als doppelt so viele Worttypen (670) als ein durchschnittlicher GdP-Text (303). Ein Worttyp wiederholt sich im Referenztext des Innenministeriums durchschnittlich 1,7-mal, im Referenztext der Gewerkschaft 2,06-mal. Während im Falle der Gewerkschaft dieses Ergebnis deutlich über dem Mittelwert aller „virgin“-Texte (1,7) liegt, befindet sich der Wert für die „Info 110“ unterhalb der für die „virgin“-Texte beobachteten Marke von 1,8.

Die Aggregation konnte die Qualität des gewerkschaftlichen Referenztextes verbessern. Auf Arbeitgeberseite konnte wegen der sehr eingeschränkten Auswahl an Texten keine Erhöhung der Wiederholungsrate einzelner Worttypen erreicht werden.

Empirische Ergebnisse

„Word Scores“ und räumliche Positionen von Worttypen

Nachdem der Text-Korpus abgegrenzt und die Referenztexte identifiziert wurden, kann die Auswertung vorgenommen werden. Im ersten Schritt wird die „ideologische“ Ausrichtung der im Datensatz auftauchenden Worttypen bestimmt. Dies erfolgt, wie schon angedeutet, anhand des Vergleiches der Referenztexte. Nur die Worttypen, die mindestens einmal in mindestens einem Referenztext auftauchen, können im politischen Raum positioniert werden, sind also „scorable“ bzw. positionierbar. Die relativen Häufigkeiten der Begriffe in „virgin“-Texten, für die die „Word Scores“ berechnet werden können, bestimmen die Positionen der untersuchten Dokumente im politischen Raum. Eine der Fragen, die in Bezug auf die Übertragbarkeit der Methode auf die Diskussion über den Reformprozess des öffentlichen Dienstes gestellt werden kann, betrifft die Möglichkeit der Identifizierung einer ausreichenden Anzahl an relevanten und positionierbaren Worttypen. Als solche verstehen wir die Begriffe, die sowohl in Referenztexten auftauchen, was ihre anfängliche Gewichtung ermöglicht, als auch in den „virgin“-Texten gebraucht werden. Die Existenz der relevanten, vergleichbaren Worttypen in Parteiprogrammen ist mit hoher Wahrscheinlichkeit vorhanden, da alle politischen Gruppierungen zu gleichen Problemen Stellung beziehen müssen. Es ist aber unklar, ob sehr unter-

schiedliche Akteure, wie Gewerkschaften und Arbeitgebervertretung auch gleiche Fragen in ihren Dokumenten ansprechen. Dies wird in diesem Abschnitt der Analyse überprüft.

Positionierbar sind in unserem empirischen Beispiel 1132 Worttypen (16 % des Text-Korpus). Ein kleiner Teil dieser Gruppe, die aus 121 Worttypen besteht, bildet eine Schnittmenge beider Referenztexte. Wir können zwar positionierbare Begriffe politisch einordnen, ihre Brauchbarkeit kann aber eingeschränkt sein. Dies trifft insbesondere auf Worttypen zu, die zwar in Referenztexten, aber in keinem anderen Text zu finden sind. Im untersuchten Datensatz werden 469 positionierbare Worttypen nur in den Referenztexten gebraucht. Wenn diese Einschränkung berücksichtigt wird, muss die Anzahl der Worttypen, die für die Positionierung von „virgin“-Texten genutzt werden können, auf 663 (10 % des Text-Korpus) korrigiert werden.

Für die Überprüfbarkeit der Qualität der Referenztexte wollen wir das folgende Gütemaß verwenden.⁸ Dieses berechnet sich auf folgende Weise: Die Anzahl der tatsächlich für die Analyse nützlichen Worttypen, also derjenigen, die positionierbar sind ($A = 1132$) abzüglich der Worttypen, die positionierbar sind, aber nur in Referenztexten auftauchen ($D = 469$) kann mit der Gesamtzahl der im Text-Korpus mehrmals benutzten Worttypen ($B = 2611$) abzüglich der Worttypen, die mehrmals, aber nur in einem Text auftauchen ($C = 215$) verglichen werden. Das Verhältnis dieser beiden Größen kann als Maß für die Güte der Referenztexte genommen werden. Das Gütemaß „Q“ würde dann wie folgt berechnet:

$$Q = \frac{A - D}{B - C}$$

Die Werte des Index Q liegen im Bereich von 0 und 1. Dies lässt sich folgendermaßen begründen.

Erstens können die Werte nicht kleiner als 0 sein, weil D immer kleiner gleich A und C immer kleiner gleich B ist. Im Falle von D ist dies dadurch zu begründen, dass alle Worttypen, die positionierbar sind, in Referenztexten auftauchen müssen. Es existiert also kein Begriff, der lediglich in Referenztexten auftaucht und nicht positionierbar wäre. Der Term C ist wiederum immer kleiner gleich B, weil jeder Worttypus, der mehrmals in einem Text benutzt wird, zwangsläufig ein mehrmals benutzter Worttypus ist.

Zweitens wird der Index Q niemals größere Werte als 1 annehmen, weil A-D immer eine Teilmenge von B-C sein wird. Es ist zwar vorstellbar, dass bei sehr umfangreichen Referenztexten A den größeren Wert haben wird als B-C, dann wird aber dieser durch D

⁸ Dieses Maß wurde von Kamil Marcinkiewicz im Rahmen der vorgestellten Analyse entwickelt.

korrigiert. Alle Begriffe, die nur in Referenztexten auftauchen werden abgezogen. A-D wird aus diesem Grund nur aus den Worttypen bestehen, die zwei Bedingungen auf einmal erfüllen: sie sind positionierbar und werden mindestens in einem Referenztext und in einem „virgin“-Text gebraucht. Der Abzug von C dient der Entfernung der nicht-relevanten Begriffe. Nach dem Abzug von C sind die Werte des Zählers und des Nenners miteinander vergleichbar. Alle Worttypen, die im Referenztext und in mindestens einem anderen Dokument auftauchen, werden mit allen mehrmals in unterschiedlichen Texten benutzen Worttypen verglichen.

In einem sehr unwahrscheinlichen Fall, wenn alle mehrmals gebrauchte Begriffe nur in jeweils einer Quelle gebraucht werden (oder es keine Begriffe die überhaupt mehrmals in Texten auftauchen gibt: $B=0$ und $C=0$), ist der Wert des Nenners gleich 0 und somit der Wert von Q unbestimmt. Dies würde darauf hinweisen, dass die Texte, die für die Analyse gewählt wurden, inhaltlich keine Gemeinsamkeiten haben.

Der Wert von Q gleich 1 würde bedeuten, dass die Referenztexte alle Worttypen beinhalten, die mehrmals in unterschiedlichen Texten auftauchen. In einem solchem Idealfall wären alle Begriffe, die sich potenziell für den Vergleich eignen könnten, erfasst. Obwohl der Wert von 1 schwer zu erreichen ist, ist es zu empfehlen die Maximierung von Q anzustreben. Das von uns vorgeschlagene Gütemaß würde in der vorgestellten Analyse 0,28 (663/2396) betragen. Unsere Referenztexte erfassen also mehr als ein Viertel aller Begriffe, die potentiell überhaupt eine Vergleichbarkeit zwischen den einzelnen Dokumenten erlauben könnten.

Ein Worttyp taucht im Datensatz durchschnittlich 4,6-mal auf. Seltener gebrauchte Worttypen können eine hohe Gewichtung aufweisen, wenn sich die Häufigkeiten ihrer Verwendung durch die unterschiedlichen Akteure unterscheiden. Dies trifft zum Beispiel auf den Begriff „Bürger“ zu. Er taucht dreimal im Referenztext der GdP auf und ist im Referenztext des Innenministeriums nicht vertreten. Der Worttyp „Bürger“ erhält somit die maximale Gewichtung mit dem Wert 10. Theoretisch gesehen ist also die Wahrscheinlichkeit, dass ein Text der diesen Begriff beinhaltet der Position der Gewerkschaft zuzuordnen ist, nahe 1. Tatsächlich taucht der Eintrag viermal in „virgin“-Texten des Innenministeriums und 23-mal in „virgin“-Texten der Gewerkschaften auf.

Im Gegensatz zu „Bürger“ ist „die“ ein sehr häufig auftauchender Worttyp, der im Kontext der Texte keine inhaltliche Bedeutung hat und daher neutral auf die Positionierung der Texte wirkt. Der 1353-mal im Datensatz verwendete Artikel macht 6,53% aller „scorable“ Wörter aus. Einträge, die sehr oft im Text-Korpus gefunden werden, werden normalerweise mit niedrigen Faktoren gewichtet, da sie in allen Referenztexten mit ähnlichen Häufigkeiten vertreten sind. Dies trifft auch auf „die“ zu. Er wird 42-mal im

Referenztext des IMB und 55-mal im Referenztext der GdP verwendet.⁹ Diese Werte werden dann für die Berechnung von posteriori-Wahrscheinlichkeiten benutzt, die wiederum angeben, wie hoch die Chance ist, dass wir den Text einer bestimmten Kategorie (bei uns entweder des IMB oder der GdP) lesen, falls wir in einem unbekannten Text („virgin“-Text) auf einen bestimmten Begriff stoßen (Lowe 2008: 361). Im Falle des Worttyps „die“ beträgt die Wahrscheinlichkeit, dass ein Text, der ihn beinhaltet, ein IMB-Dokument ist, 51,8 % und 48,2 %, dass es sich um ein GdP-Dokument handelt. Zusammenfassend können wir mit Hilfe dieses Worttyps nicht sagen, wo sich ein untersuchter „virgin“-Text im politischen Raum befindet. Die anderen Ausdrücke, die mit gewisser Sicherheit als neutral bezeichnet werden können, sind „und“, „der“ und „mit“ (*Abbildung 1*).

Die Brauchbarkeit von Worttypen, die vom Programm als „ideologisch“ eingestuft, das heißt mit extremen Werten gewichtet werden, muss zusätzlich untersucht werden. Wegen ihres seltenen Auftretens im Datensatz ist bei ihrer Interpretation Vorsicht angebracht (vgl. eine spezifische Dreieck-Struktur der Daten in *Abbildung 2*). Diese Worttypen werden als charakteristisch für eine politische Position anhand eines Vergleichs mit den Referenztexten eingestuft. Eine zusätzliche Analyse ihrer Verteilung im Text-Corpus kann also als eine Cross-Validierung der Referenzdokumente gesehen werden.

Als Beispiele solcher „ideologischer“ Begriffe, die für die Bestimmung der Positionen von „virgin“-Texten benutzt werden, können für das Innenministerium Ausdrücke wie „Reform“, „Ergebnis“, „Arbeit“, „neuen“ genannt werden. Ihre Identifizierung mit der Arbeitgeberseite scheint bei „Reform“ und „Ergebnis“ auch inhaltlich nachvollziehbar zu sein. Im Falle von „Arbeit“ und „neuen“ ist dies nicht so eindeutig. Das Gleiche lässt sich auch für zwei weitere Begriffe „schon“ und „hier“ feststellen. Alle sechs Worttypen tauchen im zweiten „Pyramidendiagramm“¹⁰ auf, das den Zusammenhang zwischen den logarithmierten Häufigkeiten und Positionen von Worttypen („Word Scores“) abbildet. Sie sind am linken Rande, etwa in der Mitte der y-Achse zu finden (*Abbildung 2 und 3*). Sie gehören somit zu den am häufigsten im Text-Korpus auftauchenden Begriffen, die eine hohe Gewichtung haben.

Eine detaillierte Analyse der Verteilung der sechs oben genannten Worttypen bestätigt im Grunde ihre Anwendbarkeit für die Unterscheidung zwischen den Texten der

⁹ Die Berechnung von „Word Scores“ zeigt, wie wichtig die Unterscheidung zwischen Worttypen und Wörtern ist. Für die Berechnung werden die Häufigkeiten von Worttypen benutzt.

¹⁰ Die Pyramidendiagramme weisen Ähnlichkeiten mit den „Eiffel Tower of Words“-Grafiken von Slapin und Proksch (2008: 715) auf, die die Parameter aus der „Wordfish“-Analyse abbilden: „word-fixed effects“, die abhängig von der Häufigkeit des Auftauchens des Worttyps sind und „word specific weights“, die Gewichte der Worttypen. Auf den Achsen des Pyramidendiagramms werden die Häufigkeit des Auftauchens und die Gewichtung eines Begriffs („Score“) dargestellt.

Gewerkschaft und des Innenministeriums. Dies ist besonders klar im Falle von „Arbeit“, „neuen“, „schon“ und „hier“ (*Abbildung 4*). Auf der Seite der Arbeitnehmer ist die Auswahl von „ideologisch“ geprägten und relativ häufig auftauchenden Begriffen einfacher, was an der größeren Anzahl der GdP-Kommentare liegt. Zu dieser Gruppe zählen: „Gewerkschaften“, „Beschäftigten“, „Bürger“, „Forderungen“, „unsere“ und „zu“ (*Abbildung 5*). Alle mit Ausnahme von „zu“ können eindeutig mit der Arbeitnehmervertretung assoziiert werden. „Zu“ ist zwar auf den ersten Blick ein neutrales Wort, es ist aber zugleich in den Ausdrücken zu finden, die sich auf die Reform des öffentlichen Dienstes beziehen, wie „zu zahm“, „zu düster“, „zu hoch“.

Bei allen oben genannten Begriffen mit Ausnahme von „zu“ weichen die Modalwerte der Verteilungen von IMB-Dokumenten und GdP-Dokumenten nicht voneinander ab. Sie liegen in der Regel bei null. Die Dominanz der Nullwerte kann als charakteristisch für die statistische Struktur der Sprache gesehen werden, in der ähnlicher Inhalt auf unterschiedliche Weise ausgedrückt werden kann. Trotz gleicher Modalwerte sind in der Grafik Unterschiede zwischen den einzelnen Balken der Histogramme feststellbar. Im Falle von Worttypen, die im Zuge der Analyse als charakteristisch für die GdP bestimmt werden, ist die Dichte in der Modalkategorie, d. h. bei der Häufigkeit gleich null, deutlich niedriger als in den Dokumenten des Innenministeriums. Als extremes Beispiel kann der Begriff „Forderungen“ gelten, der kein einziges Mal von den Arbeitgebern gebraucht wird. Trotz der mehrdeutigen Anwendbarkeit des Worttypes „zu“, ist dies hier wegen der relativen Häufigkeit, mit der dieser Worttyp im Datensatz benutzt wird besonders interessant. Dies ermöglicht das grafische Erfassen von Unterschieden in der Verteilung und nähert sich dem idealen Fall von zwei Verteilungsfunktionen mit unterschiedlichen Modalwerten an.

Die offensichtliche Schwäche der Analyse liegt in der geringen Anzahl von geeigneten Kommentaren des Innenministeriums, was eine genaue Rekonstruktion der Worttypverteilungen auf der Arbeitgeberseite verhindert. Die Histogramme des IMB sind, insbesondere wenn sich absolute Häufigkeiten zwischen den Texten voneinander unterscheiden, unterbrochen und schwierig zu interpretieren. Das zweite Problem ist die Tendenz, dass einzelne Worttypen relativ selten im Datensatz genutzt werden. Wenn der Modus bei null zu finden ist, dann können sich keine zwei unimodalen Verteilungen mit unterschiedlichen Verteilungsgipfeln, die einfach interpretierbar wären, entwickeln. Dies ist aber ein immanentes Merkmal der Sprache, das durch die Anwendung des „stemmer“, d.h. durch die Entfernung der Flexionsendungen abgeschwächt werden könnte.¹¹ Unsere Analyse der „Word Scores“ lässt sich mit der Reflexion zur Übertragbarkeit der Methode abschließen. Die Überprüfung der Relevanz von positionierbaren Begriffen kann als erfolgreich zusammengefasst werden. Es werden inhaltliche Über-

schneidungen zwischen den Referenztexten und „virgin“-Texten festgestellt. Die Referenztexte erfassen mehr als ein Viertel aller Begriffe, die sich potentiell für den Vergleich eignen. Gleichzeitig aber ist anhand der Verteilungen von einzelnen Begriffen in den Dokumenten beider Akteure zu bemerken, dass sie sich nicht immer optimal für die Unterscheidung zwischen den gegensätzlichen Positionen eignen. Die Lösung dieses Problems ist unserer Meinung nach in der Ausweitung der Auswertung auf weitere Quellen zu suchen.

„Text Scores“ und zeitliche Veränderung von politischen Positionen

Wenn die „Word Scores“, d. h. die Gewichte der einzelnen Worttypen feststehen, können die „Text Scores“ für die „virgin“-Texte berechnet werden. Dies erfolgt in „Word-scores“ wieder in zwei Phasen. Zuerst werden die „rohen“ Werte erzeugt, die im Vergleich zu den Referenztexten sehr dicht beieinander liegen. Diese relative Nähe von „virgin“-Texten zueinander ist eine Konsequenz des Skalierungsverfahrens, das auf der Gewichtung der relativen Häufigkeiten des Auftauchens eines Worttyps in neuen Dokumenten mit früher ermittelten „Word Scores“ basiert. Um eine Vergleichbarkeit mit den Referenztexten zu gewährleisten, müssen die Textpositionen im zweiten Schritt angepasst werden („rescaling“).

Die beiden Referenztexte aus dem Jahre 2002 können als Ausgangspunkt einer Zeitreihe gesehen werden. Auf den Grafiken werden ihre „ideologischen“ Positionen als zwei horizontale Linien gezeichnet, die auf der Höhe von -10 und 10 verlaufen. Nach der Anpassung der „Text Scores“ werden die Positionen der „virgin“-Texte im Koordinatensystem eingetragen. Die Umrechnung kann auf zweierlei Weise erfolgen. Laver, Benoit und Garry (2003) schlagen eine Formel vor (LBG), die die Varianz der „virgin“-Texte der Varianz zwischen den Referenztexten anpasst. Ihre Methode wurde von Martin und Vanberg (2008a) kritisiert, da sie das Ziel der Vergleichbarkeit zwischen den bekannten und zu positionierenden Dokumenten verfehlt. Die Anwendung der LBG-Transformation führt dazu, dass die Textpositionen nicht mehr zwischen -10 und 10, sondern zwischen -42,71 und 21,08 streuen. Konstant bleibt aber die Standardabweichung, die sowohl im Falle von Referenz-Texten als auch von „virgin“-Texten 14,14 beträgt.

Martin und Vanberg (2008a) empfehlen eine Lösung, die das Problem der beschränkten Vergleichbarkeit zwischen den Referenz- und „virgin“-Texten besser als die LBG-Transformation eingrenzen soll. Ihre Methode (MV) gleicht zwar die Varianzen beider Teilmengen nicht an, zwingt aber die neu berechneten „Text Scores“ innerhalb des durch die Gewichte der Referenztexte vorgegebenen Rahmens zu bleiben. Die Positionen neuer Dokumente bleiben dann zwischen den beiden in der Grafik eingefügten horizontalen Linien. Es ändert sich die Streuung, die für die „virgin“-Texte nur 4,43

beträgt (*Abbildung 6*). Die erzwungene Spannweite beeinflusst dabei den Verlauf der Positionspfade nicht, die Datenstruktur bleibt erhalten.

Eine durch die Veröffentlichung von Martin und Vanberg ausgelöste Debatte konzentrierte sich auf Vor- und Nachteile beider Verfahren. Benoit und Laver (2008: 109) empfehlen in ihrer Antwort in erster Linie auf die „rohen“ „Text Scores“ zu achten. Martin und Vanberg (2008a: 112) argumentieren auf der anderen Seite, dass ihre Methode besser die „rohen“ Positionen widerspiegelt. Sie lehnen auch die Annahme der gleichen Varianz in „virgin“- und Referenztexten als unrealistisch ab (ebd.). Selbst Benoit und Laver (2008) weisen auf einige Schwächen ihres eigenen Ansatzes hin, die in seiner Sensibilität gegenüber der Qualität der „virgin“-Texte liegen. Sie empfehlen die Anwendung der LBG-Transformation, wenn mehr als zwei hochwertige Referenztexte und eine umfangreiche, unverzerrte Stichprobe von „virgin“-Texten vorhanden sind (ebd.: 110). Beides kann zwar im von uns untersuchten Fall nicht garantiert werden, aber auch die MV-Transformation bietet lediglich auf den ersten Blick eine einfache Antwort auf die Frage nach der Vergleichbarkeit zwischen den Referenz- und „virgin“-Texten. Das einzige, was sie ermöglicht, ist der Vergleich mit den extern bestimmten Gewichten, nicht mit den Positionen, die die Referenztexte bekommen würden, wenn sie als „virgin“-Texte behandelt würden (Benoit und Laver 2008: 103).

Dabei darf nicht vergessen werden, dass es sich sowohl beim Ansatz von Laver, Benoit und Garry als auch von Martin und Vanberg um lineare Transformationen handelt, die in gleicher Weise Entwicklungstendenzen, Annäherungen und Entfernungen zwischen den untersuchten Akteuren abbilden können. Die direkte Vergleichbarkeit zwischen den Referenz- und „virgin“-Texten ist jedoch in beiden Fällen problematisch. Daher tragen wir die Positionen der Referenztexte in den Zeitreihendiagrammen lediglich als Referenzlinien ein. Sie dienen zwar als Bezugspunkte, mit Hilfe derer die „virgin“-Texte im Policy-Raum eingeordnet werden, aber die beiden Textklassen sind getrennt zu betrachten, da die Gewichtung der Referenztexte immer noch extern gesetzt ist und diese Texte selbst nicht Teil der Analyse sind. Es gibt keine einwandfrei richtige Methode für den Vergleich der beiden Textklassen. Da das Verfahren nach Martin und Vanberg für unsere Analyse nur einen geringen Mehrwert aufweist, geben wir auf unseren graphischen Darstellungen die in der Literatur häufiger gebrauchten LBG-Werte an.

Inhaltlich lässt sich sagen, dass Phasen der Annäherung und Entfernung zwischen den beiden Positionspfaden festgestellt werden können. Die Darstellung mit dem „Lewess-Smoother“ (Andersen 2009), vereinfacht die Interpretation des Verlaufes beider Kurven (*Abbildung 7*). Sowohl im Falle der Gewerkschaft als auch der Arbeitgeber wird eine leicht u-förmige Struktur beobachtet. Nach der anfänglichen Annäherung, die bis etwa 2007 andauerte, folgte eine leichte Korrektur des Kurses, die sich durch die Entfernungen der Positionen des Innenministeriums ausdrückt. Grundsätzlich kann von

einer Mäßigung der GdP-Positionen gesprochen werden, die sich jedoch ab 2007 wieder leicht von der IMB-Position entfernt. Auf der Arbeitgeberseite sind die Zeichen mehrdeutig. Während 2008 und 2009 die extremsten Positionen zu beobachten sind, folgt im April 2010 ein Text, der große Übereinstimmung mit der Position des Referenztextes der GdP aufweist.

Diese Entwicklungen legen die Vermutung nahe, dass die Gewerkschaft zunächst weiterhin kooperationsbereit war und die Verwaltungsreformen mittrug. Möglicherweise hat diese Unterstützung aber aufgrund der Enttäuschung über die Wirksamkeit der Beteiligung am Reformprozess – wie in 2.3 beschrieben – dazu geführt, dass sich die Gewerkschaft – zumindest rhetorisch – auf eine extremere Position zurückgezogen hat. Die Annäherung der Arbeitgeberseite 2010 könnte als ein Entgegenkommen und Reaktion darauf interpretiert werden, da diese unbestritten auf die Kooperationsbereitschaft der Beschäftigten für die Verwirklichung ihrer Reformziele angewiesen ist. Diese Interpretation ist jedoch im späteren Forschungsprozess mit Experteninterviews und durch zusätzliche Positionspapiere der Akteure zu validieren.

Um die anhand des Datensatzes mit individuellen Texten erzeugten Ergebnisse intern zu validieren, wurde eine zusätzliche „Wordscores“-Analyse durchgeführt. Diesmal wurden die einzelnen Dokumente beider Akteure jahresweise zusammengefügt. Anstatt für einzelne Texte, wurden die Positionen aller für ein bestimmtes Jahr in Frage kommenden Kommentare ermittelt (*Abbildung 8*). Als Anker für die Untersuchung dienen wie vorher die kumulierten Referenztexte aus dem Jahre 2002. Ein Vorteil dieses Verfahrens ist die bessere Übersichtlichkeit der Darstellung und die Angleichung der Anzahl von ermittelten Positionen bei beiden untersuchten Akteuren. In sechs von acht Zeitpunkten, die sich im untersuchten Zeitraum befinden, ist die Schätzung sowohl für die GdP als auch für das brandenburgische Innenministerium möglich. Zugleich ist aber nicht zu vergessen, dass sich durch das jährliche Zusammenfügen die Ungleichgewichte in der Länge der untersuchten Texteinheiten vergrößern. Dies liegt an der unterschiedlichen Anzahl der Kommentare pro Jahrgang auf Gewerkschafts- und Arbeitgeberseite. Während ein Text der Gewerkschaften im Durchschnitt aus 2901 Wörtern besteht, liegt der Mittelwert für die Worthäufigkeit auf der Arbeitgeberseite nur bei 837.

Die Entwicklungspfade der Positionen der Gewerkschaft und des Arbeitgebers aus den jährlich kumulierten Texteinheiten stimmen grundsätzlich mit dem Verlauf in der „Lowess-Smoother“-Darstellung überein. Besonders deutlich ist es auf der Arbeitgeberseite, wo einer Phase der Entfernung von den Gewerkschaftspositionen (2007-2009) eine Phase der Annäherung nach 2009 folgt. Auf der Seite der GdP wurde die anfängliche Politik fortgesetzt, wobei insgesamt wieder eine gewisse Milderung des Kurses mit leichtem Rückzug auf anfängliche Positionen zum Ende hin festgestellt wird.

ZUR ÜBERTRAGBARKEIT DES VERFAHRENS UND AUSBLICK

Die Analyse hat gezeigt, dass in der Literatur beschriebene Phänomene in unseren Ergebnissen wiederzufinden sind. Dies spricht für eine Übertragbarkeit des „Wordscores“-Verfahrens auf die Textart der Kommentare, trotz ihrer relativen Kürze und einer relativ schwach ausgeprägten inhaltlichen Kontroverse. Die Übereinstimmung mit der Literatur legt die Wirksamkeit des Verfahrens auch in Bezug auf den von uns untersuchten Kontext nahe.

Allerdings muss beachtet werden, dass die in der Literatur beschriebenen Phänomene auf kommunaler Ebene beobachtet wurden und diese nicht ohne weiteres auf die Länderebene übertragen werden können. Zudem muss die Interpretation auch für den hier untersuchten Sektor – die Polizei – spezifiziert werden. Allgemein darf nicht vergessen werden, dass die Positionierung der Akteure keine Aussage über ihre Zustimmung bzw. Ablehnung der Reformen zulässt, sondern lediglich ihre relativen Positionen innerhalb der Dimension der Zustimmung bzw. Ablehnung der Verwaltungsreformen und damit der Veränderung des Typus der öffentlichen Beschäftigung wiedergibt. Im weiteren Projektverlauf sollen daher die Ergebnisse mithilfe von Experteninterviews, beispielsweise mit Personalräten oder Vertretern der Gewerkschaften und Arbeitgeberverbände, sowie mit weiteren Dokumenten, in denen die Akteure ihre Positionen kundtun und die bisher nicht in die Analyse einbezogen wurden, interpretiert und validiert werden. Dies kann zudem auch noch einmal Aufschluss über die Setzung der Referenzpunkte geben.

Des Weiteren ist die interne Validität der Ergebnisse durch die Wiederholung der Analyse mit unterschiedlichen Referenztexten nach Vorbild von Klemmensen, Hobolt und Hansen (2007) zu überprüfen. Insbesondere kann die Analyse zeitlich umgedreht und mit Referenztexten aus dem Jahr 2010 durchgeführt werden, da die Auswertung der Literatur darauf hindeutet, dass sich die Positionen von Gewerkschaften und Arbeitgeberverbänden zu diesem Zeitpunkt stärker voneinander unterscheiden als im Jahre 1990 bzw. 2002 (Bogumil et al. 2008; Frieß et al. 2000; Kißler, Graf und Wiechmann 2000). Wie bereits angesprochen, erschwert die dünne Datenlage insbesondere auf Arbeitgeberseite und die ungleiche Länge der Referenz- und „virgin“-Texte die Analyse. Wir stehen daher vor einem Zielkonflikt zwischen der Maximierung der Texthomogenität und der Erreichung eines ausreichend großen Textkorpus. Es ist daher zu untersuchen, welchen Einfluss die Homogenität der Untersuchungstexte tatsächlich auf das Ergebnis hat, indem der Effekt der Inklusion von öffentlichen Stellungnahmen wie Reden und Pressemitteilungen der Akteure geprüft wird. Dies könnte auch die Fokussierung der Analyse auf den Teilaspekt des Wandels des Beschäftigungsregimes, der durch entsprechende Verwaltungsreformen ausgelöst wurde, erleichtern. Als weitere Möglichkeit, die Wiederholungsrate von Worttypen und somit die Menge an „scorable“ Worttypen zu erhöhen, kann die Anwendung des bereits erwähnten „stemmer“ angesehen werden.

Insgesamt lässt sich jedoch festhalten, dass die Ergebnisse darauf hindeuten, dass das Verfahren der quantitativen Inhaltsanalyse auf den Aushandlungsprozess der beiden Interessengruppen im Rahmen des Neuen Steuerungsmodells grundsätzlich anwendbar ist. Der Zweifel, dass die Texte aufgrund mangelnder Vergleichbarkeit in ihrer Länge sowie mangelnder inhaltlicher Kontroversität keine interpretierbaren Ergebnisse erbringen, konnte ausgeräumt werden. Die gewonnenen Ergebnisse bieten den Ausgangspunkt für weiterführende Interpretationen.

Abbildung 3: Zusammenhang zwischen Log-Häufigkeiten und „ideologischen“ Gewichten einzelner Worttypen (Auswahl extrem gewichteter und häufig verwendeter Worttypen)

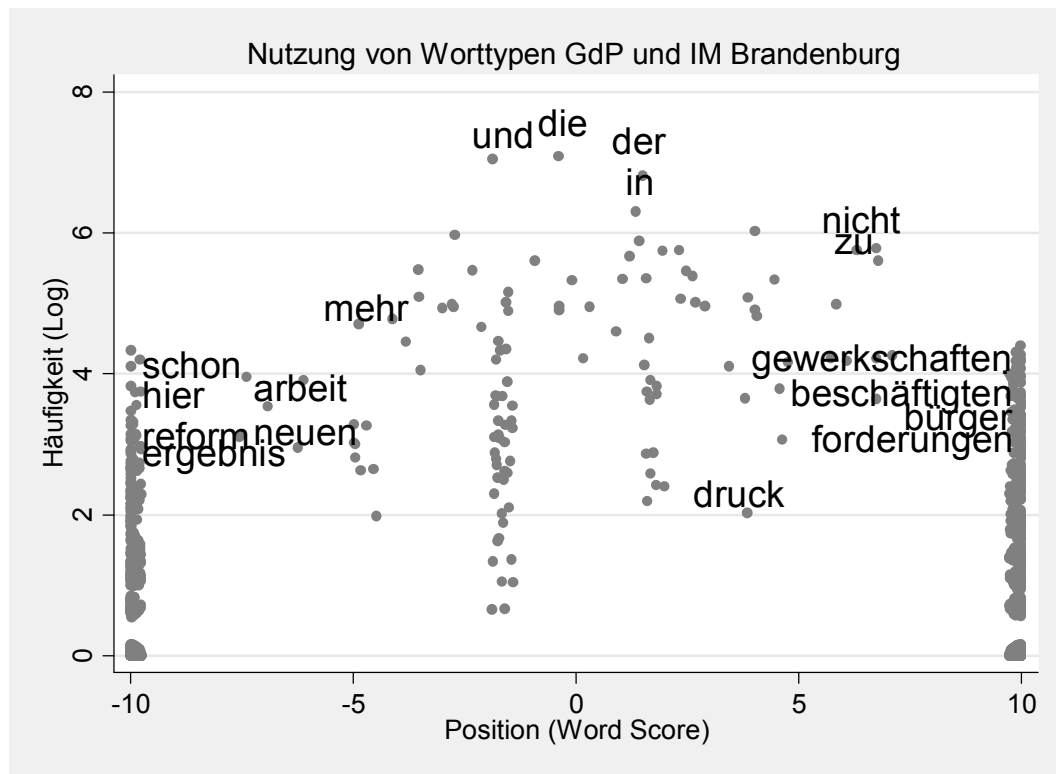


Abbildung 4: Verteilung der für das IMB charakteristischen Worttypen

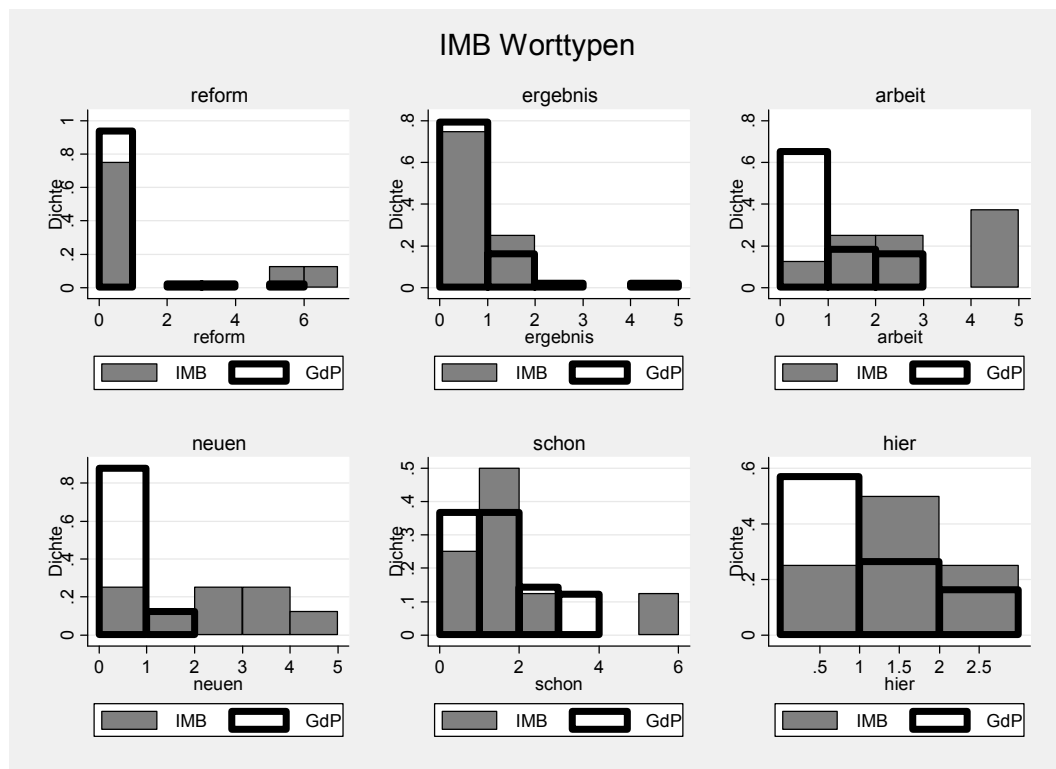


Abbildung 5: Verteilung der für die GdP charakteristischen Worttypen

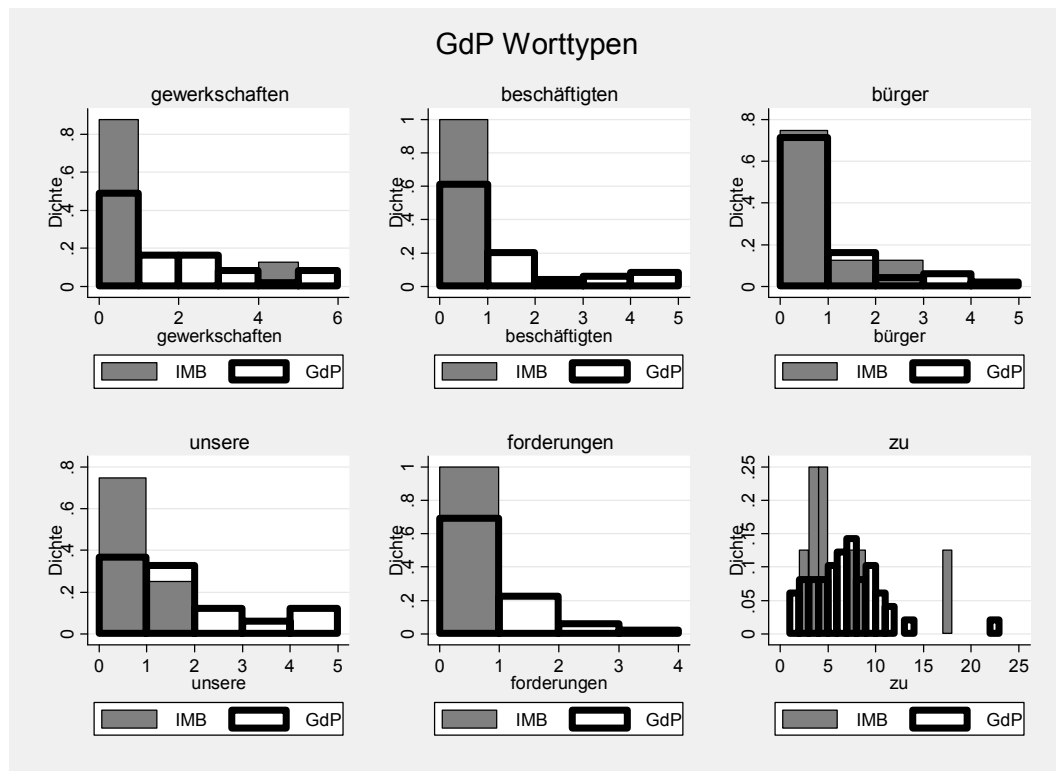


Abbildung 6: Positionen der einzelnen Dokumente nach der MV-Transformation

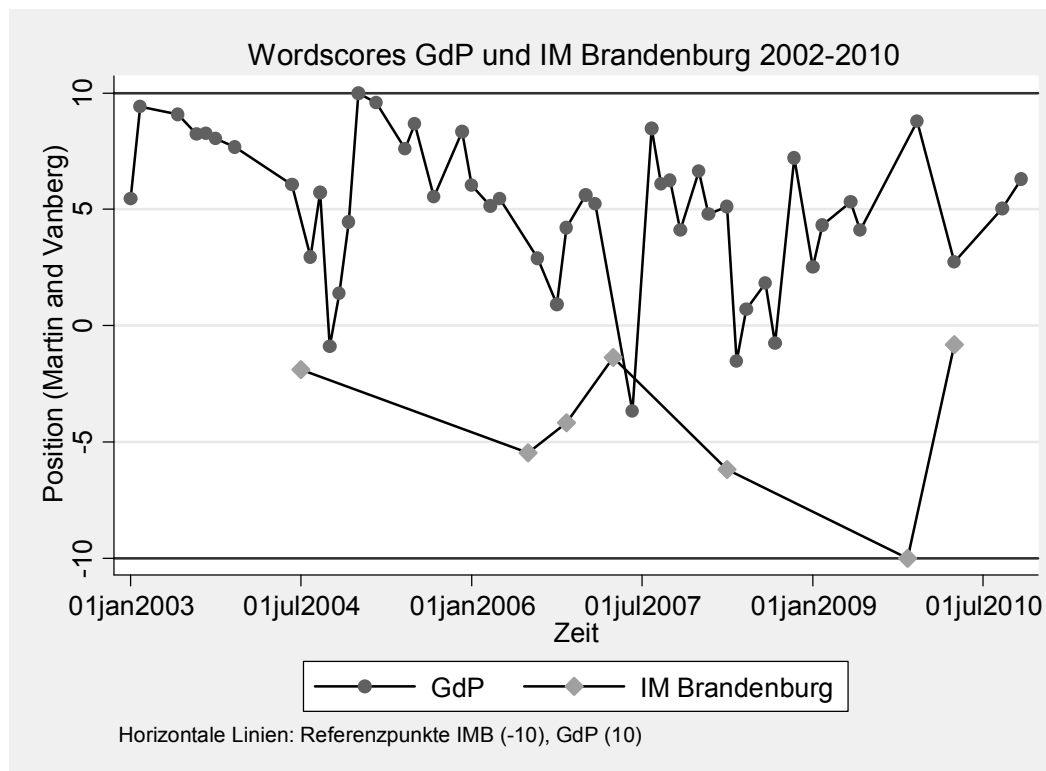


Abbildung 7: Lowess-Smoother mit MV-transformierten Werten

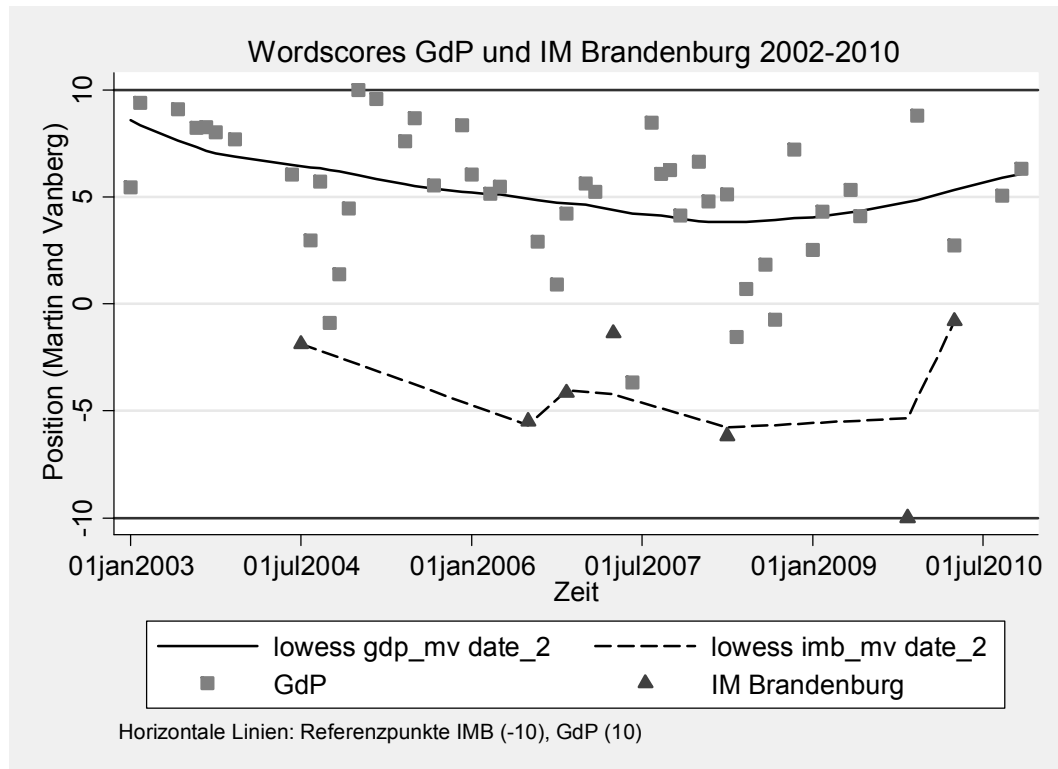
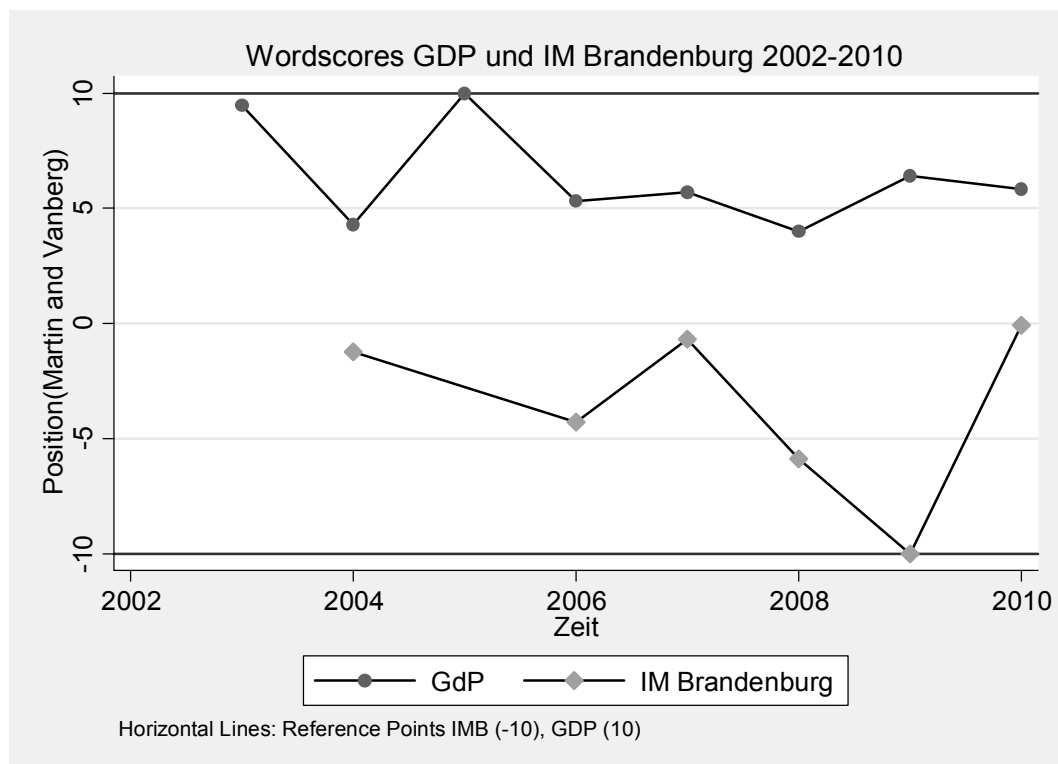


Abbildung 8: Positionen der jahresweise aggregierten Dokumente (MV-Transformation)



LITERATUR

- Albright, Jeremy J. (2010) The multidimensional nature of party competition, *Party Politics* 16(6), 699-719.
- Andersen, R. (2009) Nonparametric Methods for Modeling Nonlinearity in Regression Analysis *Annual Review of Sociology*, 67-85.
- Barthel, Christian (2008) New Public Management - symbolischer Ausdruck der Elitenzirkulation im höheren Vollzugsdienst der Polizei, in Andrea Müller Rudolf Fisch, Dieter Beck (Hg.) *Veränderungen in Organisationen – Stand und Perspektiven*: VS Verlag für Sozialwissenschaften, 377-413.
- Benoit, Kenneth und Michael Laver (2006) *Party Policy in Modern Democracies*, London: Routledge.
- Benoit, Kenneth und Michael Laver (2008) Compared to What? A Comment on ‘A Robust Transformation Procedure for Interpreting Political Text’ by Martin and Vanberg, *Political Analysis* 16(1), 101-11.
- Benoit, Kenneth, Michael Laver und Slava Mikhaylov (2009) Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions, *American Journal of Political Science* 53(2), 495-513.
- Bernauer, Julian und Thomas Bräuninger (2009) Intra-Party Preference Heterogeneity and Faction Membership in the 15th German Bundestag: A Computational Text Analysis of Parliamentary Speeches, *German Politics* 18(3), 385-402.
- Black, Duncan (1958) *The Theory of Committees and Elections*, Cambridge: At the University Press.
- Bogumil, Jörg et al. (Hg.) (2008) *Zehn Jahre Neues Steuerungsmodell – Eine Bilanz kommunaler Verwaltungsmodernisierung*, Berlin: edition sigma.
- Bortz, Jürgen (2005) *Statistik für Human- und Sozialwissenschaftler*, Heidelberg: Springer.
- Budge, Ian et al. (2001) *Mapping Policy Preferences. Estimates for Parties, Electors, and Governments 1945-1998*, Oxford: Oxford University Press.
- Castles, Francis G. und Peter Mair (1984) Left Right Political Scales - Some Expert Judgments, *European Journal of Political Research* 12(1), 73-88.
- Coffe, Hilde und Rebecca Plassa (2010) Party policy position of Die Linke: A continuation of the PDS?, *Party Politics* 16(6), 721-35.
- Coppedge, Michael (1997) A Classification of Latin American Political Parties, *Working Paper 244*, Kellogg Institute at the University of Notre Dame.
- Downs, Anthony (1957) *An Economic Theory of Democracy*, Boston: Addison Wesley.
- Ezrow, Lawrence et al. (2010) Mean voter representation and partisan constituency representation: Do parties respond to the mean voter position or to their supporters?, *Party Politics* 17(3), 275-301.
- Frevel, Bernhard und Philipp Kuschewski (2009) Police Organization and Police Reform in Germany: The Case of North Rhine-Westphalia, *German Policy Studies* 5(2), 49-89.
- Frieß, Sieglinde et al. (Hg.) (2000) *Modernisierung des öffentlichen Dienstes – eine Zukunftsbilanz*: edition der Hans Böckler Stiftung.

- Gabel, Matthew J. und John D. Huber (2000) Putting parties in their place: Inferring party left-right ideological positions from party manifestos data, *American Journal of Political Science* 44(1), 94-103.
- Gottschall, Karin und Bernhard Kittel (2007) Der Wandel des Staates als Arbeitgeber: Öffentliche Beschäftigungsregimes zwischen Effizienz und Effektivität, 1980-2010, in Stephan Leibfried et al. (Hg.) *Staatlichkeit im Wandel. Förderantrag des Sonderforschungsbereichs 597*, Bremen: Universität Bremen.
- Hils, Sylvia und Sebastian Streb (2010) Vom Staatsdiener zum Dienstleister? Veränderungen öffentlicher Beschäftigungssysteme in Deutschland, Großbritannien, Frankreich und Schweden, *TranState Working Papers* (111).
- Hinich, Melvin und Michael C. Munger (1997) *Analytical Politics*, Cambridge: Cambridge University Press.
- Hotelling, Harold (1929) Stability in Competition, *The Economic Journal* 153(39), 41-57.
- Huber, John D. und Ronald Inglehart (1995) Expert Interpretations of Party Space and Party Locations in 42 Societies, *Party Politics* 1(1), 73-111.
- Kißler, Leo, Melanie Graf und Elke Wiechmann (Hg.) (2000) *Nachhaltige Partizipation. Beschäftigtenbeteiligung als Beitrag für mehr Chancengleichheit*, Berlin: edition sigma.
- Klemmensen, Robert, Sara B. Hobolt und Martin E. Hansen (2007) Estimating policy positions using political texts: An evaluation of the Wordscores approach, *Electoral Studies* 26(4), 746-55.
- Klingemann, Hans-Dieter et al. (2006) *Mapping Policy Preferences II: Estimates for Parties, Electors and Governments in Central and Eastern Europe, European Union and OECD 1990-2003*, Oxford: Oxford University Press.
- Klüver, Heike (2009) Measuring Interest Group Influence Using Quantitative Text Analysis, *European Union Politics* 10(4), 535-49.
- Krippendorff, Klaus (2004) *Content Analysis: An Introduction to Its Methodology*, Thousand Oaks: Sage Publications.
- Laver, Michael und Kenneth Benoit (2002) Locating TDs in Policy Spaces: Wordscoring Dáil Speeches, *Irish Political Studies* 17(1), 59-73.
- Laver, Michael, Kenneth Benoit und John Garry (2003) Extracting policy positions from political texts using words as data, *American Political Science Review* 97(2), 311-31.
- Lowe, Will (2008) Understanding Wordscores, *Political Analysis* 16(4), 356-71.
- Lowe, Will (2010) *An Introduction to Austin*: University of Maastricht.
- Martin, Lanny W. und Georg Vanberg (2008a) Reply to Benoit and Laver, *Political Analysis* 16(1), 112-4.
- Martin, Lanny W. und Georg Vanberg (2008b) A robust transformation procedure for interpreting political text, *Political Analysis* 16(1), 93-100.
- Neuendorf, Kimberly A. (2002) *The Content Analysis Guidebook*, Thousand Oaks: Sage Publications.
- Nowag, Werner (2005) Kommentar/Glosse, in Siegfried Weischenberg, Hans J. Kleinsteuber und Bernhard Pörksen (Hg.) *Handbuch Journalismus und Medien*, Konstanz: UVK-Verlagsgesellschaft, 148-59.

- Proksch, Sven-Oliver und Jonathan B. Slapin (2009) How to Avoid Pitfalls in Statistical Analysis of Political Texts: The Case of Germany, *German Politics* 18(3), 323-44.
- Schmidt, Dieter (2005) Die Rolle des Arbeitskreises Methodik/Koordination im Veränderungsprozess der Polizei des Landes Nordrhein-Westfalen, in Jochen Christe-Zeyse (Hg.) *Modernes Management in der deutschen Polizei - Erfahrungsberichte aus der Praxis*, Frankfurt: Verlag für Polizeiwissenschaft.
- Schneider, Karsten (2011) Beschäftigtenbeteiligung und Mitbestimmung, in Bernhard Blanke et al. (Hg.) *Handbuch zur Verwaltungsreform*, Wiesbaden: VS Verlag für Sozialwissenschaften.
- Semmler, Jörg und Göttrik Weber (2005) Neues Steuerungsmodell, in Bernhard Blanke et al. (Hg.) *Handbuch zur Verwaltungsreform*, 3., völlig überarbeitete und erweiterte Auflage, Wiesbaden: VS Verlag für Sozialwissenschaften.
- Shepsle, Kenneth A. (2010) *Analyzing Politics. Rationality, Behavior, and Institutions* 2nd ed., New York: W.W. Norton & Company.
- Slapin, Jonathan B. und Sven-Oliver Proksch (2008) A Scaling Model for Estimating Time-Series Party Positions from Texts, *American Journal of Political Science* 52(3), 705-22.
- Smithies, Arthur (1941) Optimum Location in Spatial Competition, *Journal of Political Economy* 49(3), 423-39.
- Sperling, Hans Joachim (1998) Verwaltungsmodernisierung: Chancen für Beteiligung?, *Industrielle Beziehungen* 5(3), 322-39.
- Wiesehomeier, Nina und Kenneth Benoit (2009) Presidents, Parties, and Policy Competition, *Journal of Politics* 71(4), 1435-47.

BIOGRAPHISCHE ANMERKUNG

Kamil Marcinkiewicz und **Jennie Auffenberg** sind Wissenschaftliche Mitarbeiter an der Universität Oldenburg und am Sonderforschungsbereich *Staatlichkeit im Wandel*, Universität Bremen.

Kamil Marcinkiewicz

Telefon: +49 441 798 4795

E-Mail: kamil.marcinkiewicz@uni-oldenburg.de

Jennie Auffenberg

Telefon: +49 441 798 2974

E-Mail: jennie.marie.auffenberg@uni-oldenburg.de

Fax: +49 441 798 4855

Anschrift: Universität Oldenburg, Zentrum für Methoden der Sozialwissenschaften, Ammerländer Heerstr. 114-116, D 26129 Oldenburg

Bernhard Kittel ist Professor für Wirtschaftssoziologie an der Universität Wien und Projektleiter im Sonderforschungsbereich *Staatlichkeit im Wandel*, Universität Bremen.

Telefon: +43 1 4277 38311

E-Mail: bernhard.kittel@univie.ac.at

Fax: +43 1 4277 38318

Anschrift: Universität Wien, Institut für Wirtschaftssoziologie, Brünner Straße 72, A 1210 Wien